

HUMBERTO CERESER IBAÑEZ

**GEOMEDICINA: SISTEMA DE VISUALIZAÇÃO DE FATORES  
AMBIENTAIS E DOENÇAS EM MAPAS NA INTERNET**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre. Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, Universidade Federal do Paraná.

Orientador: Prof. Dr. Hélio Pedrini

Co-orientador: Prof. Dr. Bonald Cavalcante Figueiredo

CURITIBA

2007

## SUMÁRIO

<b>LISTA DE ABREVIATURAS E SIGLAS</b>	<b>iv</b>
<b>LISTAS DE FIGURAS</b>	<b>vii</b>
<b>RESUMO</b>	<b>viii</b>
<b>1 INTRODUÇÃO</b>	<b>1</b>
1.1 Caracterização do Problema . . . . .	1
1.2 Motivações . . . . .	3
1.3 Objetivos Geral e Específicos do GeoMedicina . . . . .	4
1.4 Contribuições . . . . .	5
1.5 Organização do Trabalho . . . . .	6
<b>2 REVISÃO BIBLIOGRÁFICA</b>	<b>8</b>
2.1 Projeto de Software . . . . .	8
2.2 Projeto com Reuso . . . . .	9
2.3 Desenvolvimento Incremental . . . . .	11
2.3.1 Prototipação Evolucionária . . . . .	12
2.4 Interface com o Usuário . . . . .	13
2.5 Visualização Geográfica . . . . .	15
2.6 Métodos Estatísticos Espaciais . . . . .	19
2.7 Modelagem de Dados Espaciais . . . . .	20
2.7.1 Organização de Dados Espaciais . . . . .	21
2.7.2 Índices de Base de Dados . . . . .	24
2.8 Padrões de Interoperação . . . . .	26
2.8.1 Protocolos de Comunicação dos Serviços <i>Web</i> . . . . .	29
2.8.1.1 Protocolo SOAP . . . . .	29
2.8.1.2 Gramática WSDL . . . . .	30

2.8.1.3	Outros Protocolos dos Serviços Web . . . . .	31
2.9	Mapas Colaborativos na Internet . . . . .	32
2.10	Proveniência e Fluxo dos Dados . . . . .	32
2.11	Sistemas Georreferenciados Aplicados à Saúde . . . . .	34
2.12	Sistemas de <i>Web Mapping</i> para Computação Geoespacial . . . . .	42
<b>3</b>	<b>PROJETO E EVOLUÇÃO DO SISTEMA</b>	<b>46</b>
3.1	Escopo do Sistema . . . . .	47
3.2	Cuidados com os Riscos do Projeto . . . . .	47
3.3	Requisitos da Base de Dados . . . . .	48
3.3.1	Dados Geoquímicos . . . . .	49
3.3.2	Dados do Sistema de Saúde . . . . .	49
3.3.3	Dados Cartográficos . . . . .	50
3.4	Requisitos da Interface com o Usuário . . . . .	51
3.5	Protótipo Inicial - API Ka-map . . . . .	51
3.5.1	Servidor HTTP Apache . . . . .	52
3.5.2	SGBD PostgreSQL e PostGIS . . . . .	53
3.5.3	UMN-MapServer . . . . .	54
3.5.4	PHP e Scripting . . . . .	55
3.5.5	AJAX, JavaScript e XML . . . . .	55
3.5.5.1	Cascading Style Sheets - CSS . . . . .	57
3.5.5.2	Document Object Model - DOM . . . . .	57
3.5.6	Escolha dos Componentes de Software . . . . .	58
3.5.7	Instalação e Configuração do Protótipo Inicial . . . . .	59
3.5.8	Avaliação do Protótipo . . . . .	60
3.6	Projeto de Evolução e Manutenção da Aplicação . . . . .	61
3.6.1	Rede de tarefas . . . . .	63
3.6.2	Diagrama Entidade-Relacionamento . . . . .	64
3.6.3	Criação e Inserção na Base de Dados Espacial . . . . .	64

<b>4</b>	<b>RESULTADOS OBTIDOS</b>	<b>66</b>
4.1	Interface do Protótipo Inicial . . . . .	66
4.2	Interface do Segundo Protótipo . . . . .	70
4.3	Interface do Terceiro Protótipo . . . . .	73
4.4	Avaliação do Protótipo . . . . .	79
<b>5</b>	<b>DISCUSSÃO SOBRE O TRABALHO</b>	<b>84</b>
5.1	Discussão sobre os Resultados . . . . .	84
5.2	Trabalhos Futuros . . . . .	86
<b>6</b>	<b>CONCLUSÕES</b>	<b>89</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>91</b>
	<b>APÊNDICE A INSTALAÇÃO E CONFIGURAÇÃO DOS PACOTES DEBIAN</b>	<b>103</b>
A.1	Instalação dos Pacotes Debian . . . . .	103
A.2	Configuração do Servidor <i>Web</i> e Implantação do <i>Framework</i> . . . . .	105
	<b>APÊNDICE B CRIAÇÃO DA BASE DE DADOS</b>	<b>106</b>
B.1	Criação da Base de Dados no PostGIS . . . . .	106
B.2	Criação de Tabelas a partir de ShapeFile . . . . .	107
B.3	Criação de Tabelas a partir de Arquivo DBF . . . . .	107
B.4	Preenchimento de Tabelas a partir de Arquivos CSV . . . . .	110
B.4.1	Preenchimento de Tabela com Número de Casos de Doença . . . . .	111
B.4.2	Preenchimento de Tabela com Teores Geoquímicos . . . . .	111
	<b>APÊNDICE C CONFIGURAÇÃO DO MAPSERVER</b>	<b>113</b>
C.1	Formatação da Entrada do MapServer - Camadas do Mapa . . . . .	113
C.2	Formatação da Saída do MapServer - Imagem Digital . . . . .	116

## LISTA DE ABREVIATURAS E SIGLAS

AJAX	<i>Asynchronous Javascript and XML</i>	Javascript e XML assíncronos
API	<i>Application Programming Interface</i>	Interface de Programação de Aplicação
ASCII	<i>American Standard Code for Information Interchange</i>	Código Americano Padrão para Intercâmbio de Informação
CGI	<i>Common Gateway Interface</i>	Interface de Entrada (Saída) Comum
CSS	<i>Cascading Style Sheet</i>	Folha de Estilo em Cascata
ER	<i>Entity-relationship</i>	Entidade-relacionamento
ESRI	<i>Environmental Systems Research Institute</i>	Instituto de Pesquisa em Sistemas Ambientais
FOSS4G	<i>Free Open Source Software for Geoinformatics</i>	Software de Código Aberto e Livre para Geoinformática
GML	<i>Geographic Markup Language</i>	Linguagem de Marcação Geográfica
HTML	<i>Hypertext Markup Language</i>	Linguagem de Marcação de Hipertexto
HTTP	<i>Hypertext Transfer Protocol</i>	Protocolo de Transferência de Hipertexto
ISO	<i>International Organization for Standardization</i>	Organização Internacional para Padronizações
MIME	<i>Multipurpose Internet Mail Extensions</i>	Extensões de Multipropósito para Correio Eletrônico
OGC	<i>Open Geospatial Consortium</i>	Consórcio Geoespacial Aberto
PHP	<i>PHP: Hypertext Preprocessor</i>	PHP: Processador de Hipertexto
RPC	<i>Remote Procedure Call</i>	Chamada de Procedimento Remoto
SCSI	<i>Small Computer System Interface</i>	Interface Pequena de Sistemas Computacional
SFS	<i>Simple Features Specification</i>	Especificação Simples de Características (geográficas)

SGBD	<i>Database Management System</i>	Sistema Gerenciador de Bases de Dados
SIG	<i>Geographic Information System</i>	Sistema de Informações Geográficas
SOAP	<i>Simple Object Access Protocol</i>	Protocolo Simples de Acesso a Objetos
SPRING	<i>Georeferencing Information System</i>	Sistema de Processamento de Informações Georreferenciadas
UDDI	<i>Universal Description, Discovery, and Integration</i>	Descrição, Descoberta e Integração Universais
UMN	<i>University of Minnesota</i>	Universidade de Minnesota
WCS	<i>Web Coverage Service</i>	Serviço Web de Cobertura
WFS	<i>Web Features Service</i>	Serviço Web de Características (geográficas)
WMS	<i>Web Map Service</i>	Serviço Web de Mapas
WSDL	<i>Web Services Description Language</i>	Linguagem de Descrição de Serviços de Internet
XML	<i>Extensible Markup Language</i>	Linguagem de Marcação Extensível
XSD	<i>XML Schema Definition</i>	Definição de Esquema XML

## LISTA DE FIGURAS

2.1	Desenvolvimento de um novo membro da família de aplicação. . . . .	10
2.2	Interfaces de um componente de software no projeto com reuso. . . . .	10
2.3	O modelo espiral. . . . .	14
2.4	Variáveis visuais e seus modos de implantação, segundo Slocum [99] - Fonte: Ramos [21], p. 24. . . . .	18
2.5	Serviços de acesso a mapas na Internet. Fonte: <i>PCI Geomatics</i> [75]. . . . .	28
2.6	Mapa empregado na pesquisa epidêmica do cólera na cidade de Londres em 1854 - Fonte: CSISS [12] . . . . .	35
2.7	Mapa da mortalidade por câncer em negros, homens e mulheres, nos Estados Unidos, entre 1970 e 1994. Fonte: <i>Atlas of Cancer Mortality in the United States</i> [27, 63]. . . . .	37
2.8	Mapa da mortalidade por câncer em brancos, homens e mulheres, nos Estados Unidos, entre 1970 e 1994. Fonte: <i>Atlas of Cancer Mortality in the United States</i> [27, 63]. . . . .	38
2.9	Arquitetura em camadas para FOSS4G em plataforma <i>Web</i> . . . . .	43
3.1	Arquitetura inicial do protótipo GeoMedicina. . . . .	52
3.2	Rede de tarefas do projeto de evolução do protótipo. . . . .	63
3.3	Diagrama entidade-relacionamento do GeoMedicina. . . . .	65
4.1	Protótipo inicial do GeoMedicina. Visualização geográfica de (a) Chumbo em sedimentos e de (b) Câncer de Colo de Intestino. Fontes: MINEROPAR 1996 e DATASUS 2004. . . . .	68
4.2	Protótipo inicial do GeoMedicina. Visualização geográfica da relação entre (a) Chumbo e Câncer de Colo de Intestino; (b) ampliação de (a). Fontes: MINEROPAR 1996 e DATASUS 2004. . . . .	69

4.3	Layout da segunda versão do protótipo GeoMedicina. Mapa com os primeiros 30 mil testes de DNA realizados e mutações do gene P53 encontradas. Fonte: IPPP 2006. . . . .	71
4.4	Ampliação do mapa com número de testes de DNA realizados e mutações encontradas. Fonte: IPPP 2006. . . . .	72
4.5	Mapa com uma classificação das mutações constatadas nas Regionais de Saúde. Fonte: IPPP 2006. . . . .	73
4.6	Arquitetura atual do protótipo GeoMedicina. . . . .	74
4.7	Mapa com a distribuição espacial do elemento Bário em águas. Fonte: MINEROPAR 1996. . . . .	75
4.8	Mapa com a distribuição espacial do elemento Chumbo em sedimentos de fundo dos rios. Fonte: MINEROPAR 1996. . . . .	76
4.9	Mapa com imagem satélite de trecho da bacia do Rio Paraná. Fonte JPL-NASA 2004. . . . .	76
4.10	Mapa com imagem satélite transparente de trecho da bacia do Rio Paraná e com imagem de fundo teor do elemento Bário. Fonte JPL-NASA 2004 e MINEROPAR 1996. . . . .	77
4.11	Mapa com imagem satélite identificando o uso do solo com área de floresta, área de agricultura e área de represa. Fonte: JPL-NASA 2004. . . . .	78
4.12	Mapa com imagem satélite identificando município com área urbana predominante. Fonte: JPL-NASA 2004. . . . .	78
4.13	Mapa com a classificação da concentração de alumínio em uma região. Fonte: MINEROPAR 1996. . . . .	79
4.14	Gráfico com resultados das avaliações da satisfação dos usuários com o protótipo.	82



## RESUMO

Este trabalho descreve o emprego de *Web Mapping* em um sistema de visualização geográfica aplicado à Epidemiologia e à Geologia Médica, denominado GeoMedicina, no qual, através de uma navegação em mapas servidos via rede de computadores, epidemiologistas e ambientalistas descobrem padrões espaciais até então desconhecidos.

No GeoMedicina são apresentadas as distribuições espaciais de elementos químicos e da incidência de doenças de interesse na região de estudo. A sobreposição dessas apresentações sugere hipóteses de associações espaciais entre as doenças e os elementos químicos que, quando comprovadas pelo estudo etiológico, identificam as áreas de risco à saúde humana, o que auxilia o processo de tomada de decisão dos responsáveis pela manutenção da saúde pública.

Na composição do sistema são integrados componentes de software livre, o que proporciona economia e flexibilidade para a adaptação do protótipo inicial conforme os requisitos levantados. *Scripts* escritos nas linguagens *PHP*, *JavaScript*, *Cascading Style Sheets (CSS)* e *XML* são adaptados na evolução da interface *Web* com o usuário. Este conjunto de tecnologias é conhecido como *Asynchronous JavaScript and XML (AJAX)*, o qual proporciona uma interação rápida e dinâmica entre o cliente e o servidor.

O sistema é construído com a metodologia de reuso de software, na qual são selecionados componentes que provêem os requisitos funcionais como a interoperação, que possibilita a expansão do sistema com a agregação de dados espaciais adquiridos de serviços remotos, e a capacidade de extensão, que possibilita a inclusão de novas funcionalidades aos componentes. O SGBD PostgreSQL e sua extensão PostGIS, o servidor de mapas UMN-MapServer, o servidor HTTP Apache e a linguagem interpretada PHP são os principais componentes deste sistema.

## CAPÍTULO 1

### INTRODUÇÃO

#### 1.1 Caracterização do Problema

O sensoriamento remoto, a coleta e análise de amostras de materiais, o gerenciamento de dados espaciais e a representação e análise destes dados pelos Sistemas de Informações Geográficas (SIG) possibilitam uma visualização geográfica de fenômenos naturais, por meio da qual podem ser feitos o monitoramento e a projeção das condições de saúde da biosfera, que é o lugar na superfície terrestre onde há vida. Além disso, a quantificação de fatores ambientais e a visualização de indicadores de saúde em mapas permite que se descubram relacionamentos de causa (fatores ambientais) e efeito (morbi-mortalidade vegetal e animal).

A visualização geográfica do teor de elementos químicos e de indicadores de saúde proporciona uma visão que auxilia a pesquisa das causas das doenças e o processo de tomada de decisão para a alocação de recursos nos locais onde é necessário um maior cuidado com a saúde da população. Esta visualização é obtida com o uso de recursos computacionais de computação gráfica, de processamento de imagens digitais, de funções geoestatísticas, geodésicas e algébricas que constituem os SIG. Um SIG é um sistema de informações com capacidade de representação, modelagem, simulação e análise de dados espaciais.

A distribuição espacial heterogênea da incidência de câncer entre populações que vivem em diferentes regiões do mundo corrobora com a hipótese da influência dos fatores ambientais sobre as neoplasias. Estudos mostram que incidências de algumas neoplasias tendem a se agrupar em certas regiões onde a taxa de incidência apresenta-se notavelmente mais elevada. Um exemplo é o câncer pediátrico no córtex da glândula supra-renal, cuja incidência no Estado do Paraná é de 12 a 18 vezes superior à incidência reportada nos Estados Unidos e Europa [81].

Para a epidemiologia, a visualização da distribuição espacial das doenças em mapas permite, por meio da exploração desta visualização, a avaliação do risco em locais onde sejam confirmadas hipóteses da toxicidade dos elementos químicos. Em uma subsequente visão, as

delimitações das áreas de risco à saúde norteiam a distribuição dos recursos necessários para os cuidados com a saúde pública, nos locais onde a população é mais afetada.

Uma exploração científica requer uma interface humano computador eficiente e uma modelagem da base de dados adequada, onde os aspectos inerentes à aplicação, como o espaço-temporal, devem ser considerados para que o sistema responda a questões básicas como *quando* e *onde*. A eficiência do sistema está vinculada a uma modelagem abrangente de sua base de dados, que deve prever e organizar os possíveis relacionamentos entre os objetos do sistema. Neste trabalho, apresentam-se os procedimentos empregados na construção da base de dados, incluindo a aquisição dos dados, a modelagem da base de dados e a importação dos mesmos para esta base.

O poder de aquisição dos dados aumenta quando o sistema tem a capacidade de interoperação, através da qual o sistema se expande ao adquirir dados geográficos produzidos por serviços disponíveis na rede de computadores. Neste trabalho, mostra-se como, a partir da escolha de componentes compatíveis com padrões internacionais, esta capacidade foi obtida e como serviços de mapas foram usados para enriquecer a visualização geográfica do sistema. A construção do sistema é feita com o reuso de software, processo em que são integrados componentes prontos e testados, o que aumenta a confiabilidade do sistema e a agilidade de sua produção. A escolha de componentes de uma família de aplicação é justificada para atender a um conjunto de requisitos, sendo o principal a interoperação.

Outra necessidade imprescindível dos sistemas desta natureza, onde são tratados e relacionados dados do meio ambiente e de saúde, é a capacidade de extensão ou modificação para que alguma funcionalidade específica seja acrescentada. Esta característica é dificultada quando se empregam componentes de software fechados ou comerciais, os quais, por sua natureza, impedem que sejam feitas modificações. Por isso, adotou-se o emprego de componentes de software de código aberto e livre, os quais permitem que sejam feitas as adaptações necessárias para atender aos requisitos particulares deste sistema. Este trabalho mostra como estes componentes foram integrados na construção do sistema e quais foram as modificações realizadas para a adaptação às necessidades do sistema.

## 1.2 Motivações

A atitude do homem em relação à biosfera interfere nas condições dos biótopos, que são os locais propícios à vida, condições estas que podem ser favoráveis ou prejudiciais a sua saúde. Hoje, as consequências do efeito estufa no planeta Terra convencem a opinião pública de que são necessárias ações para conter as emissões de gases na atmosfera. Entretanto, as principais causas do aquecimento global iniciaram-se há 200 anos, no início da era industrial, quando a ciência e a tecnologia eram empregadas, principalmente, visando à produtividade por meio da exploração dos recursos naturais. Hoje, a inteligência humana, enriquecida pela teia mundial de saberes, passa a ser usada no sentido de promover a sustentabilidade e a saúde do planeta. Neste contexto, a descoberta das causas de doenças crônicas que afetam as populações consiste em um meio de garantir qualidade de vida e corrigir atitudes prejudiciais à vida.

A incidência do câncer vem aumentando nas últimas décadas no mundo. Em alguns países, o câncer já é a principal causa de morte da população. Estudos mostram que o Brasil poderá seguir esta tendência. Em 2000, o câncer era a terceira maior causa de morte por doenças, sendo fatores contribuintes o envelhecimento da população, propiciado pelo processo de urbanização e pelas ações de recuperação da saúde, que aumentam o tempo de exposição contínua aos fatores ambientais [48]. Em 2005, o câncer era a segunda maior causa de morte por doenças, sendo a maior exposição a determinados riscos ambientais apontada como uma das características responsável por uma mudança no perfil de morbi-mortalidade nos últimos 20 anos [13].

Neste contexto, conhecer a localização das doenças e de seus agentes etiológicos para prevenir o contato com elementos tóxicos e cuidar da saúde da população afetada pelos fatores de risco é um anseio de epidemiologistas e ambientalistas, responsáveis pela vigilância em saúde. Atualmente, há um consenso de que fatores ambientais freqüentemente exercem influência nos mecanismos genéticos que contribuem para o surgimento das doenças [114].

O GeoMedicina apresenta em mapas da região de interesse as seguintes informações: as classificações das diferentes taxas de incidência de doenças crônicas; as concentrações de elementos químicos no ambiente, cujas influências sobre as doenças são conhecidas e outras serão

investigadas; os relacionamentos importantes entre os elementos químicos e as doenças, constituindo hipóteses que são analisadas por pesquisadores. Na avaliação do risco, comprovando-se as hipóteses, o pesquisador pode identificar as áreas de risco à saúde humana.

Na apresentação destes resultados é empregada a metodologia de visualização geográfica, derivada da visualização científica que, segundo Ramos [21], “pode ser considerada como o uso de ferramentas computacionais com o propósito de aprimorar a análise de dados, tornando evidentes informações que não poderiam ser percebidas de outra forma, ou tornando mais rápido o processo de percepção. Dessa forma, a visualização permite que se veja o que não poderia ser visto”. Esta metodologia surgiu a partir da semiologia gráfica, criada por Bertin [9], a qual procura estabelecer formas eficientes de representação gráfica dos fenômenos naturais para facilitar a percepção cognitiva humana.

### **1.3 Objetivos Geral e Específicos do GeoMedicina**

O objetivo geral é auxiliar a pesquisa epidemiológica em neoplasia e o processo de tomada de decisão por agentes da saúde pública, por meio da apresentação de padrões em mapas digitais que delimitem as regiões consideradas como de risco, com acesso facilitado à informação digital através da Internet. Para atingir este objetivo geral, são organizados alguns objetivos específicos, descritos a seguir:

- caracterizar a distribuição espacial de mortalidade para doenças crônicas, por município na região de estudo;
- caracterizar a distribuição espacial de elementos químicos de interesse, coletados em pontos estratégicos nas bacias hidrográficas e no solo da região de estudo;
- exibir as relações importantes entre variáveis químicas e de saúde por meio de visualização geográfica;
- sugerir hipóteses de relacionamento entre doenças e fatores ambientais, contribuindo para a identificação das áreas de risco à saúde humana.

Pretende-se também que o GeoMedicina integre conhecimentos sobre a exposição aos produtos de origem natural (geológica) e/ou manipulados pelo homem, tais como mercúrio, chumbo, arsênio, manganês, bário, cádmio, resíduos agrotóxicos, benzeno, tolueno e outros produtos tóxicos, dos quais são encontrados na literatura indicação de relação causa-efeito com certas doenças. Doenças de evolução aguda, e principalmente crônica, como câncer, doenças de pele, neurológicas, dentre outras, poderão ser melhor avaliadas em um sistema integrado com um diagnóstico mais preciso do ambiente químico. O sistema terá também condições de apresentar um perfil sobre a qualidade dos mananciais, as variações naturais no quimismo da rede hidrográfica superficial e os focos de poluição industrial, agrícola e doméstica, em ambientes rurais e urbanos.

## **1.4 Contribuições**

O GeoMedicina agrega metodologias e tecnologias da Computação e da Cartografia e tem como usuários pesquisadores da Medicina, da Química e da Geologia, além de agentes da administração pública, caracterizando um projeto multidisciplinar. A interface desenvolvida, adequada aos perfis dos usuários potenciais do sistema, oferece uma interação inovadora, a qual auxilia a pesquisa das causas das doenças crônicas com a visualização de padrões antes ocultos à pesquisa, sendo assim, uma ferramenta útil aos pesquisadores e aos agentes responsáveis pelos cuidados com a saúde pública.

O emprego de softwares livres apresenta vantagens, mas também traz alguns desafios, como o conhecimento das linguagens de programação empregadas e da arquitetura do protótipo inicial. Este conhecimento é necessário para que sejam feitas adaptações no software, as quais beneficiam a comunidade de software livre mundial que consulta a documentação produzida sobre o conhecimento assimilado. No processo de software, é empregada a técnica de desenvolvimento com reuso de software, no qual um protótipo inicial integra os componentes existentes.

No projeto da interface com o usuário é empregada a tecnologia AJAX, a qual proporciona uma interação mais natural do usuário com o sistema, com o mínimo de interrupções causadas pela espera de atualizações de conteúdo, economizando o uso da banda de rede. Apenas são

veiculados pela rede os conteúdos dinâmicos, modificados após a última interação do usuário, ao invés da transmissão de todo o conteúdo, como ocorre no estilo tradicional de interação cliente-servidor de uma interface *Web* que não emprega esta abordagem.

A modelagem dos dados espaço-temporais em uma base de dados relacional prevê os relacionamentos entre elementos químicos e doenças e a representação no tempo e no espaço destas variáveis. Neste trabalho é apresentado um diagrama entidade relacionamento do esquema da base de dados. Também é fornecido o procedimento usado para importar dados não espaciais de planilhas eletrônicas para a base de dados espacial.

## 1.5 Organização do Trabalho

O Capítulo 2 apresenta uma revisão bibliográfica da metodologia e tecnologia atualmente empregadas no desenvolvimento de sistemas desta natureza, a qual serve como base para a escolha dos componentes de software integrados na construção do sistema. Aspectos sobre a visualização geográfica e modos de interação definem o projeto da interface humano-computador. A estatística espacial e a modelagem dos dados espaciais são estudadas com o objetivo de acrescentar funcionalidade ao sistema. No final do capítulo é apresentado um estudo dos requisitos de sistemas de informações geográficas aplicados com a finalidade de monitorar e descobrir as causas das doenças.

O Capítulo 3 mostra o levantamento dos requisitos do sistema, os componentes escolhidos no projeto com reuso, os quais constituem uma interface de programação de aplicação, as descrições da arquitetura da aplicação e exemplos da interface com o usuário do sistema. Também é fornecido um projeto de manutenção e evolução da aplicação, no qual é planejado um conjunto de atividades necessárias para o aprimoramento do protótipo inicial, incluindo a modelagem dos dados espaciais.

O Capítulo 4 apresenta os resultados obtidos com os protótipos, avaliados com a participação do usuário que interage com os projetistas no aprimoramento dos requisitos de software, manifestando sua impressão a respeito da interface por meio de um questionário de satisfação do usuário da interface.

O Capítulo 5 traz uma discussão sobre este trabalho e sugestões de trabalhos futuros.

As conclusões deste trabalho são apresentadas no capítulo 6. Finalmente, três apêndices são incluídos para descrever os procedimentos para instalação e configuração de pacotes de software livre utilizados no projeto, a criação da base de dados do sistema GeoMedicina e a configuração do Servidor de Mapas.



## CAPÍTULO 2

### REVISÃO BIBLIOGRÁFICA

Este capítulo apresenta uma revisão bibliográfica de tópicos de Engenharia de Software, em que são estudados modelos como: orientados a reuso, desenvolvimento incremental, desenvolvimento em espiral e prototipação evolucionária e a interação humano-computador. Uma vez que o principal componente da interface com o usuário é a imagem digital de mapas, são apresentados alguns conceitos elementares da representação cartográfica. Considerações sobre a metodologia estatística inerente ao projeto e sobre a modelagem de bases de dados espaciais e os padrões e protocolos que garantem a expansão distribuída do sistema são estudadas. Trabalhos referentes a sistemas georreferenciados aplicados à saúde são revistos, quanto à tecnologia e metodologia aplicadas e aos aspectos dos requisitos de software para o projeto de um sistema desta natureza. Características da tecnologia *Web Mapping* também são estudadas.

#### 2.1 Projeto de Software

Pressman [85] afirma que há três fases no desenvolvimento de software: definição, desenvolvimento e manutenção. Na fase de definição, a equipe de desenvolvimento identifica quais dados têm de ser processados, qual função e desempenho são desejados, quais interfaces devem ser estabelecidas, quais restrições de projeto existem e quais critérios de validação são exigidos. Esta fase é expressa em um projeto de software que representa os requisitos de software em gráficos, tabelas ou linguagens, os quais descrevem a estrutura de dados, a arquitetura, o procedimento algorítmico e as características da interface.

Em um projeto, a documentação é desenvolvida e controlada para garantir a manutenção do software. Um controle das mudanças é instituído de forma que elas possam ser aprovadas e acompanhadas. A análise de riscos de projeto possibilita a atenuação destes, o que evita o fracasso do projeto através dos seguintes passos: identificação dos riscos, avaliação dos riscos,

disposição dos riscos em ordem de prioridade, administração dos riscos, resolução dos riscos e monitoração dos riscos.

A criação de uma rede de tarefas ou rede de atividades permite que a equipe de desenvolvimento de software controle o trabalho a ser feito no tempo previsto. Responsabilidades por tarefas são atribuídas aos integrantes da equipe. A definição de uma seqüência de atividades, considerando a interdependência das tarefas, possibilita que algumas tarefas sejam executadas em paralelo e que outras essenciais sejam priorizadas de forma a garantir o progresso contínuo do processo.

## **2.2 Projeto com Reuso**

Segundo Sommerville [102], um processo de software é um conjunto de atividades e resultados associados que levam à geração de um produto de software. Esse processo pode envolver o desenvolvimento do software desde o início, embora, cada vez mais, ocorra o caso de um software novo ser desenvolvido mediante a expansão e a modificação de sistemas já existentes. Para o autor, atualmente poucos sistemas são completamente novos, o que leva a considerar o desenvolvimento e a manutenção como estágios integrados e contínuos. Ao invés de dois processos separados, é mais realista pensar na engenharia de software como um processo evolucionário, em que o software é continuamente modificado ao longo de seu tempo de duração, em resposta a requisitos em constante modificação e às necessidades do cliente. O autor ainda afirma que construir sistemas com reuso de componentes é essencial para o rápido desenvolvimento de software e para redução de custos e de riscos.

Uma das abordagens mais eficazes do reuso se baseia na noção de família de aplicação, cujos membros formam uma arquitetura de domínio específico e cujo núcleo comum é reutilizado cada vez que uma nova aplicação é construída. Nesta construção pode ser necessário o desenvolvimento de alguns componentes adicionais ou a adaptação dos componentes existentes, em uma especialização funcional, a fim de atender aos requisitos particulares do sistema que está sendo construído. Sommerville recomenda, no projeto com reuso de família de aplicação, uma arquitetura do sistema em camadas, onde o núcleo que agrega os componentes essenciais do sistema é separado em uma camada mais interna e os componentes específicos, suscetíveis

a adaptações, são dispostos em camadas superiores. A figura 2.1 mostra a produção de um novo membro por meio de adaptações feitas sobre a família escolhida.

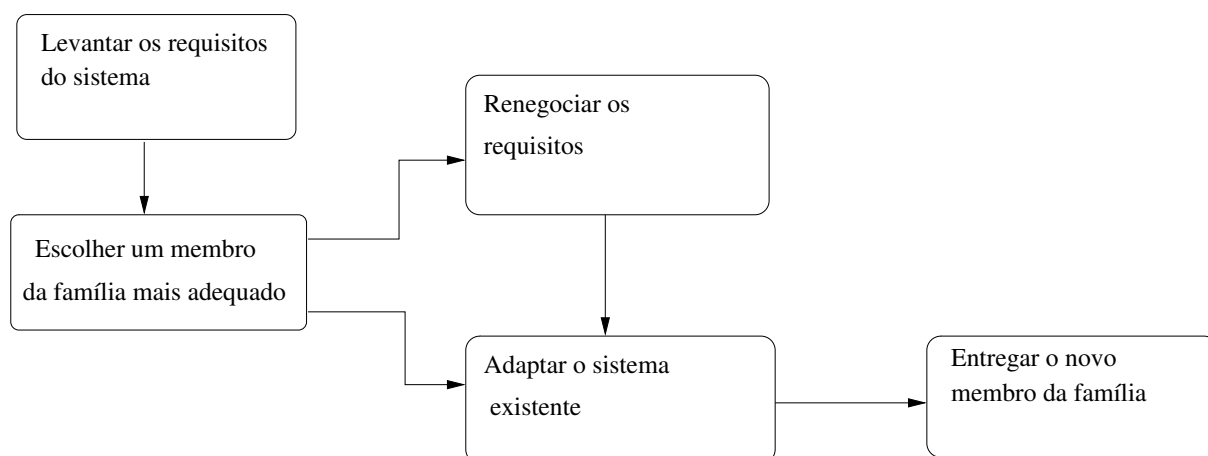


Figura 2.1: Desenvolvimento de um novo membro da família de aplicação.

No modelo orientado a reuso, os componentes são selecionados conforme os requisitos especificados. Como em geral não se encontram componentes que atendam exatamente à especificação original, os requisitos são adaptados de forma a refletir a realidade dos componentes escolhidos, ou seja, cada subsistema integrado agrega ao sistema que está sendo construído um conjunto de requisitos inerentes. Os componentes são definidos por duas interfaces: a interface REQUER, que especifica os serviços a serem fornecidos por outros componentes e a interface PROVÊ, que especifica os serviços a serem fornecidos por este componente. A figura 2.2 representa as interfaces que definem um componente reusável.

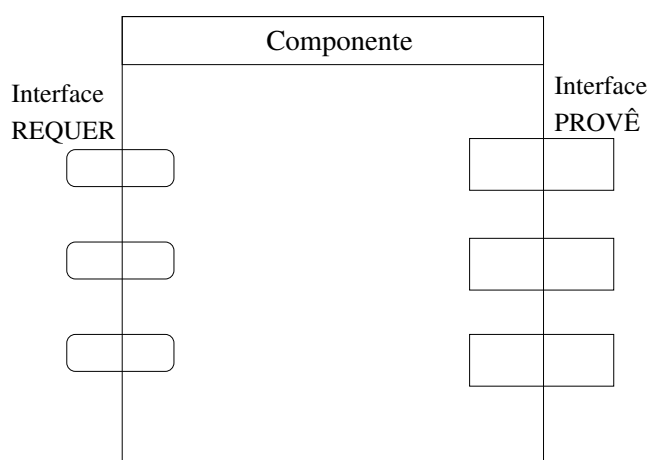


Figura 2.2: Interfaces de um componente de software no projeto com reuso.

Segundo Pressman [85], o custo de aquisição de um software existente quase sempre será menor do que o custo para desenvolver um software equivalente. Em se tratando de software livre, este custo é reduzido, pois não é exigido o pagamento de licenças de software. Entretanto, uma cautela é necessária: caso haja a necessidade de alguma modificação no software existente, às vezes, o custo para modificar um software existente é maior do que o custo para desenvolver um software equivalente.

Sommerville [102] alerta para os problemas a que está sujeita esta abordagem. As adequações nos requisitos são inevitáveis, o que pode resultar em um sistema que não atenda às reais necessidades dos usuários. Além disso, perde-se algum controle sobre a evolução do sistema uma vez que novas versões dos componentes para o reuso não estão sob o controle dos projetistas. No entanto, ao se tratar de software livre, estes problemas podem ser resolvidos uma vez que o código fonte dos componentes é aberto e poderá ser modificado pela equipe de desenvolvimento, desde que esta tenha o domínio da linguagem e da arquitetura do referido componente.

## 2.3 Desenvolvimento Incremental

A abordagem de desenvolvimento incremental consiste na entrega de produtos de software aos clientes em estágios, onde cada estágio contém um incremento que corresponde a um conjunto de funcionalidades, sendo as mais importantes incluídas nos primeiros estágios. Estes incrementos evoluem gradativamente para a versão final, em um processo que se vale das avaliações dos usuários para o aprimoramento da especificação dos requisitos e a identificação dos ajustes necessários para o aperfeiçoamento do software. Sommerville [102] apresenta diversas vantagens do desenvolvimento incremental:

- Os clientes não precisam esperar até que todo o sistema seja entregue, para então tirarem proveito dele. O primeiro estágio satisfaz seus requisitos mais importantes e, assim, o software pode ser imediatamente utilizado.
- Os clientes podem usar os primeiros incrementos como um protótipo e obter uma experiência que forneça os requisitos para estágios posteriores do sistema. Isto propicia

um comprometimento maior dos usuários com o produto obtido no final do processo.

- As funcionalidades mais importantes do sistema são testadas desde o início o que implica em uma maior robustez do sistema produzido, o qual, provavelmente, terá suas funcionalidades principais bem depuradas e isentas de falhas.

Estas vantagens do desenvolvimento incremental são apontadas aqui como condições para que o sistema seja realmente usado com sucesso e evite-se o abandono do uso pelos clientes após a sua implantação, um risco que é altamente relacionado com interfaces inadequadas ao usuário.

### **2.3.1 Prototipação Evolucionária**

Uma forma de minimizar o risco de fracasso de um projeto é fazer com que o cliente tenha uma participação mais efetiva no processo de software, onde o envolvimento do cliente é alcançado nas avaliações e validações dos incrementos do protótipo pelos usuários do sistema e diretores da organização. Este processo iterativo, no qual a interação entre clientes e a equipe de desenvolvimento é freqüente, traz benefícios apontados por Sommerville [102]:

- Um protótipo inicial é aprontado com agilidade e serve para mostrar a viabilidade e a utilidade da aplicação para a direção da organização.
- À medida que as funções do sistema são apresentadas nos protótipos, possíveis equívocos no levantamento de requisitos pela equipe de desenvolvimento ou na expectativa da funcionalidade do sistema pelos usuários podem ser identificados.
- A interação entre projetistas e usuários, nos sucessivos incrementos do sistema, faz com que haja uma maior aproximação entre o sistema e as necessidades dos usuários.

Um risco com o uso da prototipação evolucionária é o incremento de propriedades funcionais ao sistema sobre um protótipo reusado no qual haja deficiência de propriedades não funcionais essenciais, como, por exemplo, o desempenho. Uma reestruturação do protótipo para corrigir esta deficiência custaria muito e o prosseguimento do projeto nestas condições poderia levar

ao seu fracasso. Para evitar este risco, Sommerville [102] recomenda que os protótipos iniciais tenham uma estrutura sólida de modo que possam ser mantidos por muitos anos, sejam confiáveis e eficientes. Aqui a manutenibilidade prolongada é entendida como um quesito imprescindível, uma vez que a interoperação e a capacidade de expansão são essenciais para a evolução do sistema com a inclusão de novas interfaces.

A confiabilidade de um sistema é apresentada por Sommerville [102] em quatro dimensões principais: disponibilidade, confiabilidade, segurança e proteção. Disponibilidade é a capacidade de o sistema estar em funcionamento a qualquer momento em que seus serviços sejam necessários; confiabilidade é a capacidade de o sistema responder de forma correta, conforme especificado, às interações dos usuários; segurança é a característica de o sistema ser livre de falhas que causem danos às pessoas ou ao seu ambiente; e proteção é a capacidade de o sistema resistir a invasões acidentais ou deliberadas.

Um dos benefícios do projeto com reuso é a maior confiabilidade. Os componentes que são empregados no reuso, a princípio, são mais confiáveis que os componentes novos. Eles já foram experimentados e testados em diferentes ambientes. Os defeitos de projeto e de implementação são descobertos e eliminados no uso inicial dos componentes, reduzindo assim o número de falhas quando o componente é reusado. Aqui se ressalta o potencial do software livre neste aspecto, onde uma comunidade mundial, que cresce diariamente, experimenta e testa os componentes que serão reusados.

Uma forma de reduzir os riscos do projeto de software é empregar o modelo espiral, o qual exige uma consideração direta dos riscos técnicos em todas as etapas do projeto e, se adequadamente aplicado, deve reduzir os riscos antes que se tornem problemáticos. Entretanto, este modelo exige experiência na avaliação dos riscos e confia-se nessa experiência para o sucesso. A figura 2.3 mostra as etapas de planejamento para este modelo, análise de riscos, engenharia e avaliação do cliente, propostas por Pressman [85].

## **2.4 Interface com o Usuário**

Um aspecto importante para garantir a aceitação do sistema é o projeto de sua interface com o usuário, o qual, para ser eficaz, deve seguir os princípios básicos de familiaridade, consistência

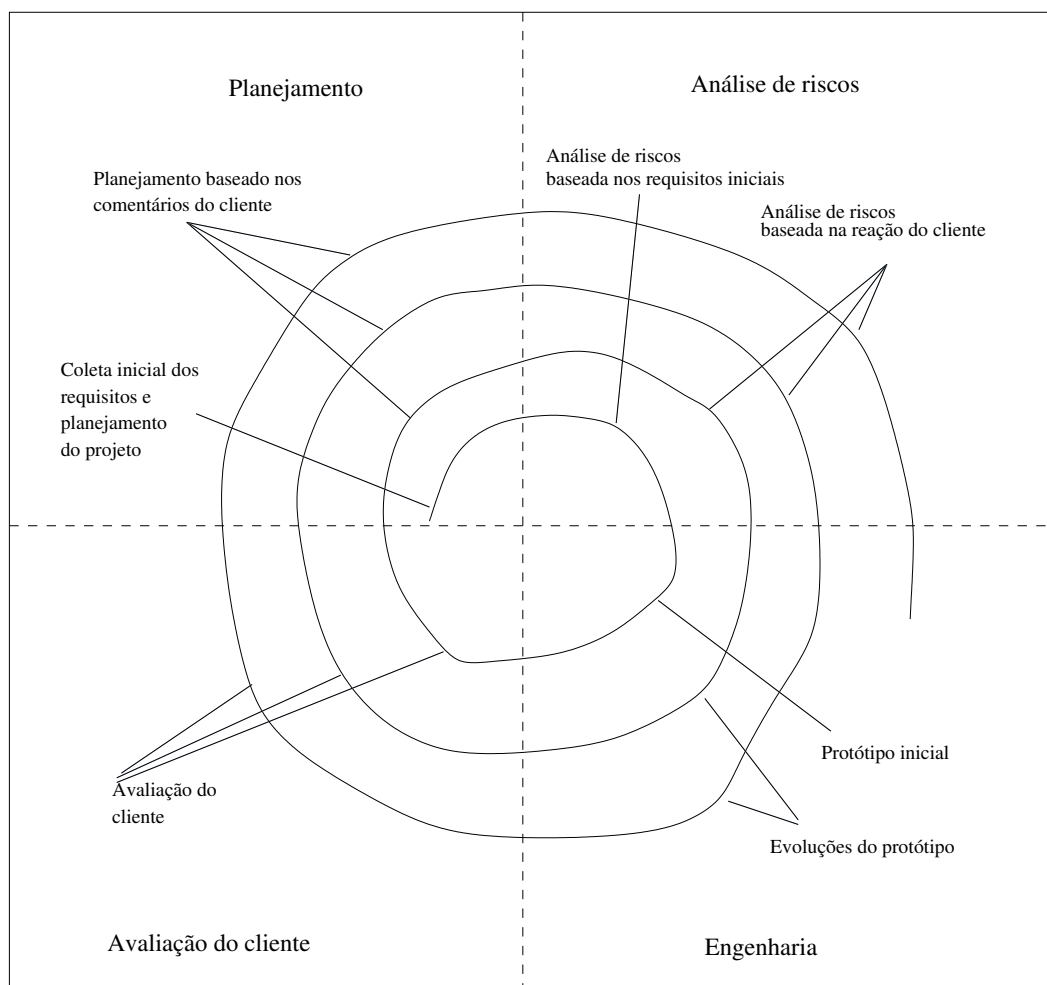


Figura 2.3: O modelo espiral.

e orientação ao usuário. A familiaridade está relacionada ao ambiente computacional a que o usuário está acostumado e a terminologia empregada nas opções de menu e rótulos. A consistência é obtida com um comportamento regular da interface com o usuário, o que simplifica o aprendizado e aumenta a confiança do usuário no sistema. A orientação ao usuário deve proporcionar a sensação de que o usuário está no controle, a qual é obtida com o emprego de realimentações (*feedbacks*) ao usuário, os quais devem esclarecê-lo sobre o que está ocorrendo após o mesmo ter realizado uma ação sobre o sistema. Isto evitará que uma seqüência de ações dos usuários cause um mal funcionamento ou corrompa o sistema.

Shneiderman [97] classificou os modos de interação em cinco estilos principais: manipulação direta, seleção de menu, preenchimento de formulários, linguagem de comando e linguagem natural.

## 2.5 Visualização Geográfica

A visualização geográfica segue o conceito da engenharia semiótica que afirma existir uma comunicação bidirecional na interação entre a interface do sistema (mapa) e o usuário (leitor). Esta comunicação é eficiente quando o sistema “percebe” as necessidades do usuário e informa a este aquilo que lhe interessa, e quando o usuário consegue comunicar o que deseja por meio dos dispositivos desta interface. Uma comunicação eficiente contribui para que haja uma exploração particular e individual dos componentes do mapa em um ambiente interativo, quando o usuário realiza suas próprias buscas e análises e, portanto, chega a um conhecimento novo [21].

Segundo Bertin [9], a representação cartográfica eficiente deve ser construída para possibilitar a visão do fenômeno representado e não sua leitura, ou seja, quanto mais natural for a apreensão do fenômeno representado, mais eficiente será a imagem gráfica.

Conceitos elementares de cartografia são a base para a confecção de qualquer mapa. O primeiro conceito a ser considerado é o modo de implantação da informação geográfica que, basicamente, pode ser pontual, linear ou areal. Uma informação é pontual (zero-dimensional) quando relacionada a um par de coordenadas, abscissa e ordenada. Este modo de implantação é empregado quando a forma do objeto geográfico pode ser desprezada. Uma informação é linear quando relacionada a uma sequência de pontos que compõem uma linha (unidimensional). Uma informação é areal (bidimensional) quando relacionada a uma sequência de pontos que formam um polígono fechado (limites políticos ou naturais) ou a um conjunto de pontos que se estendem ao longo do espaço (fenômenos físicos contínuos). Além desses três modos, Slocum [99] e outros autores propuseram o conceito de modo de implantação  $2\frac{1}{2}$ -dimensional e tridimensional. O modo  $2\frac{1}{2}$ -dimensional é implantado quando a cada coordenada está associado um único valor de um fenômeno, ou seja, cada coordenada é univariada. No modo de implantação tridimensional, a cada coordenada está associado, além do valor de uma altura acima/abaixo de um referencial, o valor de um ou mais fenômenos, ou seja, cada coordenada é bivariada ou multivariada.

Há dois tipos genéricos de mapas, os mapas de referência, que fornecem informações sobre a localização de objetos, e os mapas temáticos, que mostram a distribuição espacial de um



fenômeno [31]. Dados temáticos podem ser categorizados em qualitativos e quantitativos. Dados qualitativos diferenciam os objetos pelo tipo, espécie ou condição, enquanto dados quantitativos indicam as diferenças em quantidades ou valores associados aos objetos. Os mapas temáticos são assim denominados por conter informações acerca de um único tema [23]. Um exemplo de tema qualitativo é o de classes de uso do solo e um exemplo de tema quantitativo é o do volume de agrotóxico empregado. Em um mapa podem ser expressas informações qualitativas e quantitativas, como, por exemplo, mapas de uso do solo mostrando o volume de agrotóxico empregado, sendo, neste caso, denominados de mapas politemáticos.

Outro conceito importante para a produção eficiente de mapas temáticos é o da descrição de dados através de escalas de mensuração, as quais descrevem os dados em três níveis: nominal, ordinal e intervalar ou de razão. Dados na escala nominal diferem pelo tipo e não podem ser ordenados. Um exemplo de dados na escala nominal é o de área rural e área urbana. Dados na escala ordinal podem ser ordenados, mas possuem somente valores relativos (baixo, médio, alto), não sendo possível estabelecer uma diferença entre os mesmos uma vez que não há valores numéricos a eles associados. Dados na escala intervalar ou de razão possuem valores numéricos associados aos mesmos sendo possível classificá-los segundo estes valores numéricos.

A distribuição espacial de um fenômeno pode ser contínua ou discreta. Um fenômeno discreto é aquele que ocorre em localizações distintas e com espaço entre essas localizações onde este fenômeno está ausente. Um exemplo de fenômeno discreto é o uso do solo na agricultura, onde os tipos de culturas variam ao longo do espaço. Um fenômeno contínuo é aquele em que não há lacunas de sua ocorrência ao longo do espaço, mas somente variações. Normalmente, os fenômenos físicos apresentam uma distribuição espacial contínua, sendo um exemplo a temperatura atmosférica.

Na representação da distribuição espacial de um fenômeno, seja ele contínuo (concentração de elementos químicos) ou discreto (ocorrência de doenças em um município), é feita por meio de aplicação de variáveis visuais, as quais devem ter um emprego apropriado para que a comunicação seja eficiente. Ramos [21] afirma que Slocum, complementando os estudos de Bertin e outros, estudou a eficácia da aplicação das variáveis visuais, subdividindo-a em baixa,

média e alta. Esta classificação é dependente da escala de mensuração (nominal, ordinal ou intervalar) e do modo de implantação da informação (pontual, linear, areal, 2  $\frac{1}{2}$ -D ou 3-D) por meio das variáveis visuais apresentadas na figura 2.4.

Slocum estudou a efetividade das variáveis visuais para cada escala de mensuração em fenômenos areais. A tabela 2.1 apresenta as conclusões deste estudo.

Variáveis visuais	Escala de mensuração		
	Nominal	Ordinal	Intervalar e razão
Espaçamento	<i>B</i>	<i>M<sup>c</sup></i>	<i>M<sup>c</sup></i>
Tamanho	<i>B</i>	<i>M</i>	<i>M</i>
Altura em perspectiva	<i>B</i>	<i>M<sup>a</sup></i>	<i>A<sup>b</sup></i>
Orientação	<i>A</i>	<i>B</i>	<i>B</i>
Forma	<i>A</i>	<i>B</i>	<i>B</i>
Arranjo	<i>A</i>	<i>B</i>	<i>B</i>
Brilho	<i>B</i>	<i>A</i>	<i>M</i>
Matiz	<i>A</i>	<i>A<sup>d</sup></i>	<i>M<sup>d</sup></i>
Saturação	<i>B</i>	<i>M</i>	<i>M</i>

Tabela 2.1: Eficácia das variáveis visuais para cada escala de mensuração em fenômenos areais, segundo Slocum [99] - Legenda: (B) Baixa; (M) Média; (A) Alta; (a) Desde que diferenças com alturas sejam sugestivas de diferenças numéricas, uso com cautela para dados em escala ordinal; (b) Unidades de enumeração ocultas e ausência de orientação são problemas; (c) Não é esteticamente agradável; (d) Matizes devem ser ordenados cuidadosamente, por exemplo, amarelo, laranja, vermelho. - Fonte: Ramos [21], p. 25.

Ramos [21] afirma também que combinações destes fatores, modo de implantação, escala de mensuração e distribuição espacial, foram estudadas por Kraak e Ormeling [49]. Este estudo chegou a nove tipos de representações das quais quatro são apresentadas por apresentarem maior relevância para este trabalho. Na *representação coroplética*, cuja palavra é originada do grego *choros* (área) e *pletos* (valor), os diferentes valores são atribuídos a unidades de área. Os valores de uma série de dados a serem mapeados são submetidos a um processo de classificação e as classes obtidas são representadas em polígonos a elas associados. Valores absolutos devem ser padronizados pelo quociente com a área do polígono ou a população contida neste.

Mapas coropléticos são usados para representar fenômenos discretos. A *representação por isolinhas*, do grego *iso* que significa igual, é constituída por linhas que unem pontos de igual valor. Mapas de isolinhas são usados para representar fenômenos contínuos. Na *representação*

	Pontual	Linear	Areal	2 ½ - D	3 - D
Espaçamento					
Tamanho					
Altura em perspectiva					Não é possível
Orientação				Não recomendado	
Forma				Não recomendado	
Arranjo				Não recomendado	
Brilho					

Mapas coloridos					
Matiz					
Brilho					
Saturação					

Figura 2.4: Variáveis visuais e seus modos de implantação, segundo Slocum [99] - Fonte: Ramos [21], p. 24.

por *figuras proporcionais*, os tamanhos das figuras variam em função de valores absolutos, é usada para representar fenômenos discretos associados a pontos ou áreas. Na *representação*

*por pontos*, cada ponto está relacionado a um valor fixo do fenômeno apurado e deve ser disposto o mais próximo possível da localização exata onde ele ocorre. Uma maior ou menor concentração de pontos ao longo da região comunica a maior ou menor incidência do fenômeno mapeado.

## 2.6 Métodos Estatísticos Espaciais

A produção de mapas temáticos requer uma padronização das contagens para que se possa comparar as diferentes populações no espaço. O estimador de taxas brutas é instável quando expressa o risco de doenças raras ou quando a população sob risco é pequena. Outros possíveis estimadores incluem a regressão de Poisson, estimativas de riscos relativos e mapas de probabilidade. Porém, estes ainda não consideram a dependência espacial do fenômeno. Para considerar a dependência espacial das incidências de doenças nos municípios, podem-se usar medidas de associação espaciais tais como os índices I de Moran ou G de Geary, métodos de suavização tais como os estimadores de Bayes empírico global ou local, ou estimativas dadas por modelos com estruturas locais tais como modelos auto-regressivos condicionais, que combinam a informação do município com a de sua vizinhança [7].

Com o uso de métodos geoestatísticos é possível caracterizar os padrões de dependência espacial e, a partir destes, obter predições de interesse na região ou sub-regiões de interesse, tais como mapas de teores do atributo, zonas críticas e probabilidade de exceder valores críticos. As informações e predições sobre atributos químicos na região de estudo, integradas sob a área de cada unidade administrativa, fornecem dados a serem relacionados com os números de casos de doenças. Neste contexto, modelos de regressão, lineares ou lineares generalizados [29], cujas respostas são variáveis de doenças, investigam a existência de associações com concentrações dos elementos químicos. Tais modelos podem ser expandidos para incluir associações espaciais e incorporar a incerteza associada aos erros de predição dos atributos químicos.

Em plataforma de software de código aberto e livre, os métodos estatísticos são geralmente implementados por meio da linguagem de computação estatística R [87], a qual tem se tornado um ambiente padrão onde novas metodologias são comumente adicionadas. A linguagem R é interpretada, mas internamente procedimentos podem utilizar código compilado

para processamento rápido. O sistema é de código aberto, disponível na Internet e distribuído também em formato compilado para várias plataformas. O mecanismo de pacotes adicionais permite a incorporação rápida de novas metodologias, bem como diversas opções para interfaces com outras linguagens e aplicativos. Um exemplo, é a interface com o SGBD PostgreSQL por meio da linguagem de procedimentos PL/R [82]. Com a linguagem PL/R é possível escrever funções e *triggers* dentro do SGBD, usando a maior parte dos recursos da linguagem de computação estatística R. Pela flexibilidade, número de usuários e qualidade do projeto, o R tem se tornado a linguagem predominante para computação estatística. Além de rotinas convencionais do sistema R, pacotes adicionais a este sistema tais como SPDEP [10] e geoR [89] podem ser usados para a programação dos procedimentos de estatística espacial descritos anteriormente.

## 2.7 Modelagem de Dados Espaciais

Um modelo é uma visão abstrata de um sistema, que ignora alguns de seus detalhes, mas auxilia na estruturação e compreensão do comportamento de um sistema. Aqui, um modelo pode ser entendido como uma generalização de um sistema real baseada em computador. Um modelo amplamente usado para o projeto de aplicações com bases de dados é o modelo conceitual entidade-relacionamento (ER). Elmasri e Navathe [64] descrevem o processo de *design* da base de dados em passos simplificados expostos a seguir:

- Levantamento e análise de requisitos. A análise das entrevistas com os usuários gera um documento com um conjunto conciso de requisitos de dados. Em paralelo são também levantados os requisitos funcionais da aplicação, os quais consistem nas operações (transações) a que serão submetidos os dados.
- Criação do esquema conceitual da base de dados, onde são adicionados aos requisitos de dados uma descrição detalhada dos tipos de dados, relacionamentos e restrições. Em seguida, as transações correspondentes às operações realizadas pelos usuários são definidas. Modificações no esquema conceitual podem ser necessárias caso alguma funcionalidade não possa ser especificada.

- Implementação da base de dados em um SGBD, passo chamado de *design* lógico da base de dados, o qual resulta em um esquema da base de dados.
- *Design* físico da base de dados, onde as estruturas de armazenamento e organizações dos arquivos para a base de dados são especificadas.

### 2.7.1 Organização de Dados Espaciais

Os dados espaciais podem ser organizados segundo dois formatos principais, o formato *raster*, que consiste em imagens formadas por tabelas de *pixels* aos quais são atribuídos valores transformados em cores pelos softwares de processamento de imagens que as interpretam. Sobre estas imagens é feito um georreferenciamento com a atribuição de coordenadas geográficas a alguns *pixels* da imagem, segundo um sistema de projeção.

Um sistema de projeção é usado para projetar no plano a superfície geóide da terra. Normalmente ocorre alguma distorção no processo de projeção, sendo que os diferentes sistemas de projeção preservam algumas propriedades originais em detrimento de outras, como por exemplo, preservar ângulos e distorcer áreas.

Outro formato de dados espaciais é o formato vetorial, que consiste em primitivas gráficas, como ponto e seguimento de reta, isoladas ou combinadas, as quais são transformadas em imagens pelos softwares de computação gráfica que as interpretam. Estas primitivas gráficas são posicionadas em relação a um referencial segundo um sistema de projeção pré-definido. Dentre os formatos vetoriais, o *ShapeFile* é bastante difundido pelas empresas de produção cartográfica.

*ShapeFile* é um formato de dados espaciais definido pela ESRI [32]. Em um *ShapeFile* são armazenadas, em coordenadas vetoriais, as geometrias e a localização no espaço dos objetos geográficos, como pontos e polígonos; e os respectivos atributos descritivos, como nomes de municípios e de microrregiões. No documento *ESRI ShapeFile Technical Description* [95], pode-se verificar que um *ShapeFile* é composto por, pelo menos, três arquivos básicos, um arquivo principal (extensão *shp*) onde, em cada registro de tamanho variável, há uma descrição de um objeto geográfico com uma lista de vértices; um arquivo índice (extensão *shx*) onde

cada registro contém um *offset* do arquivo principal; e uma tabela *dbase* onde em cada registro estão contidos os atributos descritivos dos objetos geográficos.

Alguns formatos de dados proprietários tornaram-se padrões internacionais “de fato”, como é o caso do *ShapeFile*. Entretanto, Mitchell [58] adverte que muitas empresas têm criado seu próprio formato, os quais, quando adotados, restringem a liberdade do usuário na escolha da ferramenta de geoprocessamento mais adequada as suas necessidades. No Brasil, a equipe do INPE criou um formato ASCII para o SPRING [20], sendo necessário o emprego de um conversor para tornar compatíveis os principais formatos de dados de outros Sistemas de Informações Geográficas (SIG) para o SPRING. Esta abordagem possibilita o uso de outros formatos, mas limita as possibilidades de escolha do usuário aos softwares desenvolvidos pelo INPE para acesso aos formatos convertidos.

Muitos softwares de geoprocessamento, devido ao reuso de componentes no seu desenvolvimento, tornaram-se híbridos ao separar o gerenciamento da geometria (dados geoespaciais) do gerenciamento dos atributos (dados descritivos). Isto é feito para o aproveitamento da estrutura e do tratamento de dados espaciais legados, geralmente baseada em arquivos como os *ShapeFiles* ou os em ASCII, mas com a evolução de um gerenciamento dos atributos via Sistema Gerenciador de Bases de Dados (SGBD). Entretanto, esta abordagem é considerada imprópria por Rigaux [90], o qual lista suas principais desvantagens:

- A coexistência de modelos de dados heterogêneos, o que implica em dificuldades na modelagem, uso e integração.
- A perda parcial de funcionalidade básica de um SGBD, como técnicas de recuperação, pesquisa e otimização.

Por outro lado, um software de geoprocessamento que interage com uma base de dados relacional com extensão para tratamento de dados espaciais adquire a capacidade de um Sistema de Informações Geográficas. Esta extensão adiciona ao esquema relacional suporte para tipos de dados geométricos, indexação espacial e operadores geométricos. Assim, com o uso do SQL pode-se criar e manipular dados geográficos através de relacionamentos de sobreposição espacial, de predicado binário e através de mudança de sistemas de coordenadas. Uma função

de predicado binário retorna um valor booleano, indicando se dois objetos espaciais têm um determinado relacionamento, como por exemplo *intercepta* ou *está dentro*. Uma função de sobreposição espacial toma um ou mais objetos espaciais e retorna um novo, como por exemplo intersecção ou *buffer*.

Rigaux afirma que a primeira tarefa na modelagem de dados espaciais é a definição do esquema da base de dados que descreve a estrutura da informação gerenciada pela aplicação, bem como as restrições a respeito dos dados armazenados na base. Um esquema conceitual é obtido através de um modelo de entidade-relacionamento (ER), o qual descreve genericamente os objetos, chamados neste estágio de entidades, junto com os relacionamentos entre eles.

Esquemas conceituais devem ser construídos para cada tema de interesse, que é uma coleção de objetos geográficos os quais correspondem a uma entidade do mundo real. Modelos geográficos representam uma forma de descrever e manipular temas e seus objetos em um ambiente de base de dados, incluindo a capacidade topológica que consiste em um conjunto de relacionamentos espaciais existente entre objetos.

Dada a complexidade das entidades geográficas do mundo real e dos relacionamentos de composição intrínsecos existentes entre estas, é necessário definir os objetos geográficos quanto a sua atomicidade espacial (resolução geográfica) e a hierarquia entre os mesmos. Um objeto geográfico complexo consiste em outros que, por sua vez, podem ser atômicos ou complexos. A definição em notação sintática a seguir expressa as noções de objeto geográfico atômico, complexo e tema, através da qual se pode depreender que um tema é um conjunto de objetos geográficos homogêneos (ou seja, objetos que têm a mesma estrutura ou tipo).

tema = (objeto geográfico)

objeto geográfico = (atributos, geometria) // objeto atômico

                  | (atributos, objeto geográfico) // objeto complexo

Dentre os modelos para representação vetorial, estudou-se o modelo *spaghetti* [24] no qual a descrição da geometria de qualquer objeto da coleção é independente dos outros objetos, ou seja, nenhuma topologia é armazenada o que implica que todos os relacionamentos topológicos devem ser computados sob demanda. Nesta estrutura, há redundância na representação, por



exemplo, a fronteira entre duas áreas adjacentes é representada duas vezes. As vantagens deste modelo são: a habilidade de representação de objetos heterogêneos, o que possibilita a mistura de pontos, linhas e polígonos sem nenhuma restrição; e a simplicidade, obtida pela independência, o que permite a inserção de objetos pelo usuário final da aplicação.

Outro modelo também estudado foi o modelo topológico [24], no qual a topologia é armazenada, o que agiliza a computação dos relacionamentos topológicos (como a adjacência) e evita a redundância na representação mas, por outro lado, cria estruturas de dados complexas o que aumenta o custo do percurso por tais estruturas [90].

## 2.7.2 Índices de Base de Dados

Os sistemas de bases de dados são usados em sistemas de informações geográficas porque eles oferecem certos mecanismos para a organização do armazenamento e a recuperação da informação, os quais teriam de ser desenvolvidos a partir do zero em uma aplicação que usasse um sistema de arquivos simples. Uma questão importante a ser considerada é a agilidade na procura da informação constante em um campo da base de dados. Quando a quantidade de registros aumenta, uma busca seqüencial que percorre, um a um, os valores armazenados em um campo pode se tornar lenta. Para evitar isto, são criadas estruturas de dados que funcionam como índices da base de dados. Assim como o índice de um livro, um índice em base de dados evita que uma busca por uma palavra seja feita com a leitura seqüencial de todos os valores de um campo em uma tabela.

Em relação à velocidade de acesso e organização dos dados, a forma como os mesmos são armazenados em blocos de disco magnético e a leitura destes blocos pelo hardware controlado pelo sistema operacional são os principais gargalos do sistema. A função dos índices é diminuir os acessos a blocos de disco através da subdivisão de dados categorizados por algum critério de semelhança de atributos, sejam estes alfanuméricos ou espaciais. Assim, as consultas em bases de dados são aceleradas com o uso de índices, que são estruturas de acesso organizadas em uma lista de apontadores para todos os blocos de disco que contêm gravações de um determinado campo. O tamanho dos arquivos de índices é bem menor que o dos dados por ele indexados e são ordenados de forma que uma busca binária na lista seja eficiente. Além disso,

quando o índice torna-se extenso, pode-se criar um próximo nível de índice, onde o segundo nível apresenta uma quantidade menor de apontadores para o primeiro nível [64].

Um índice primário consiste em pares de valores ordenados, <chave primária, apontador para um endereço de bloco em disco>, onde o número destes pares é determinado pela quantidade de blocos em disco e os valores de chaves constantes nos pares são chamados de âncora de bloco. Um índice secundário consiste em pares de valores ordenados, <valor de campo não ordenado, apontador para um endereço de bloco em disco>. Caso o campo não ordenado seja uma chave secundária, haverá um par para cada registro. Caso o campo não ordenado não seja chave secundária, o número de pares pode ser reduzido com o uso de uma indireção, a qual é estruturada em uma lista de apontadores para campos com o mesmo valor e um primeiro nível de índice com apontadores para o primeiro elemento desta lista.

Um índice para chave secundária ou primária pode ter seu número de acesso a blocos reduzido com o emprego de um índice multinível, o qual é preferido mesmo em relação aos índices primários devido ao problema da inserção e remoção de elementos em uma lista ordenada, como é o caso da chave primária. Para resolver este problema, é deixado espaço livre em cada um dos blocos do índice multinível reservado para a inserção de novas entidades. Isto é chamado de índice multinível dinâmico e é geralmente usado usando estruturas de dados chamadas árvores  $B$  e  $B^+$ , as quais limitam a profundidade da árvore (níveis de índice) com uma configuração de árvore balanceada e com poucos espaços livres no caso de eliminação. As árvores  $B^+$  suportam somente os predicados de ordenação, menor, igual, ou maior ( $<$ ,  $=$ ,  $>$ ).

Índices espaciais têm a função de facilitar as pesquisas por dados espaciais, principalmente quando há uma grande quantidade de objetos geográficos. Um índice amplamente usado para esta finalidade é o baseado na estrutura árvore R, a qual subdivide a área da pesquisa em sub-retângulos, tantos quantos forem necessários para agrupar entidades de uma mesma classe. Estes índices suportam somente pesquisas N-dimensionais, como: contém, está contido e é igual. Devido à natureza retangular desta estrutura, estes índices espaciais somente são bons para comparações entre objetos desta forma geométrica, chamada nos SIG de retângulos envolventes (*bounding box*). Por isso, nas bases de dados espaciais é implementada uma

forma de processamento espacial em duas fases [88]:

- a primeira fase é a busca indexada com retângulos envolventes, a qual percorre todos os registros da tabela, e
- a segunda fase é o teste de processamento espacial acurado, o qual percorre um subconjunto retornado pela primeira fase.

Uma estrutura chamada de árvore de busca genérica (*Generalized Search Tree* - GiST) é usada para implementar tanto árvores B quanto árvores R. Esta estrutura extensível provê funcionalidade para todos os tipos de árvores e permite que programadores desenvolvam índices para qualquer tipo de dados, com suporte a qualquer método de busca a estes dados. Assim como a estrutura árvore B, uma GiST é balanceada e contém pares <chave, apontador>. As chaves, ao invés de serem números inteiros como nas árvores B, são membros de uma classe definida pelo programador e representam uma propriedade que é verdadeira para todos os itens de dados alcançáveis pelo apontador associado. Por exemplo, chaves em uma árvore  $B^+$  como GiST são intervalos de números (todos os dados apontados estão entre 4 e 6); chaves em uma árvore R como GiST são *bounding boxes* (todos os dados apontados estão na Califórnia) [107].

## 2.8 Padrões de Interoperação

Uma necessidade a ser considerada na concepção do sistema é a capacidade de interoperação com outros sistemas, o que possibilita a comunicação entre aplicações, embora estas tenham sido desenvolvidas de forma independente. Esta capacidade é obtida quando algum componente do sistema segue um padrão internacional para a troca de dados. Esta troca permite que temas, produzidos por outros sistemas remotos, sejam compartilhados com sistema local e vice-versa, sem que haja a necessidade de replicação de bases de dados, o que evita a cópia de grandes quantidades de dados para o computador local.

Os padrões de interoperação para Internet, denominados *Web Services* são independentes da linguagem de programação, do sistema operacional e das plataformas. Um sistema em cuja

concepção adotou-se estes padrões torna-se um sistema modular no qual se pode agregar novos produtos de softwares através da consulta a múltiplas fontes e/ou servidores de dados. Na última década duas organizações têm trabalhado em conjunto para o estabelecimento destes padrões para mapas, o *OpenGIS Consortium (OGC)* [72] e o *International Organization for Standardization / Technical Committee 211 (ISO/TC 211)* [104]. Um sistema que segue os padrões estabelecidos pelo OGC oferece vários benefícios [92], incluindo:

- Distribuição de dados geoespaciais e aplicações entre plataformas, sistemas operacionais, computadores e linguagens. Serviços OGC são neutros quanto à plataforma e à linguagem.
- Integração de funcionalidades e dados geoespaciais nas aplicações personalizadas através de rotinas implementadas em *Web Services Description Language (WSDL)* [109], as quais requerem os serviços *Web*.
- Integração de infra-estrutura que está sendo construída para a habilitação da arquitetura de Serviços *Web*, como ferramentas de desenvolvimento, servidores de aplicação, protocolos de mensagens, segurança e definições de fluxo.

O grupo ISO/TC 211 / OGC, cujo objetivo é o estabelecimento de um conjunto de padrões estruturados para informações concernentes a objetos ou fenômenos que, direta ou indiretamente, estão associados a uma localização relativa ao planeta Terra, estabeleceu o padrão ISO 19107, o qual padroniza o esquema espacial para as informações geográficas digitais. Neste esquema é definido o conceito de *features*, que são abstrações de fenômenos do mundo real com uma localização geográfica e as unidades fundamentais de informação geográfica. Dados vetoriais consistem em primitivas geométricas e topológicas usadas, separadamente ou em combinação, para a construção de objetos que expressam as *features* geográficas. A linguagem *Geographic Markup Language (GML)* - ISO 19136 [67], a qual é codificada em XML é usada para o transporte e armazenamento de informação geográfica, incluindo ambas, a geometria e as propriedades das *features* geográficas.

O OGC criou os padrões de *Web Services* para mapas conhecidos como *Web Features Service (WFS)* - ISO 19142/43 [70] e *Web Map Service (WMS)* - ISO 19128 [71]. A partir

destes padrões, uma aplicação pode acessar produtos cartográficos de servidores remotos. Os serviços de *WMS* geralmente fornecem imagens digitais processadas a partir da cobertura de satélites. Os serviços *WFS* fornecem dados de *features* no formato GML.

Um outro padrão, o *Web Coverage Service (WCS)* - ISO 19123 [69], é uma extensão do *WMS* na qual são adicionadas valores ou propriedades das localizações geográficas que as descrevem em detalhes, permitindo consultas complexas sobre os dados e fornecendo-os junto com suas semânticas originais, as quais podem ser interpretadas e extrapoladas ao invés de simplesmente reproduzidas. Além disso, este padrão retorna representações de fenômenos que variam no espaço e que relacionam o domínio espaço-temporal a um intervalo de propriedades (possível multidimensional). Um outro padrão, o *Simple Features Services (SFS)*, define como as aplicações manipulam os SGDB espaciais através do SQL. A figura 2.5 ilustra o resultado de acesso a esses serviços de mapas na Internet.

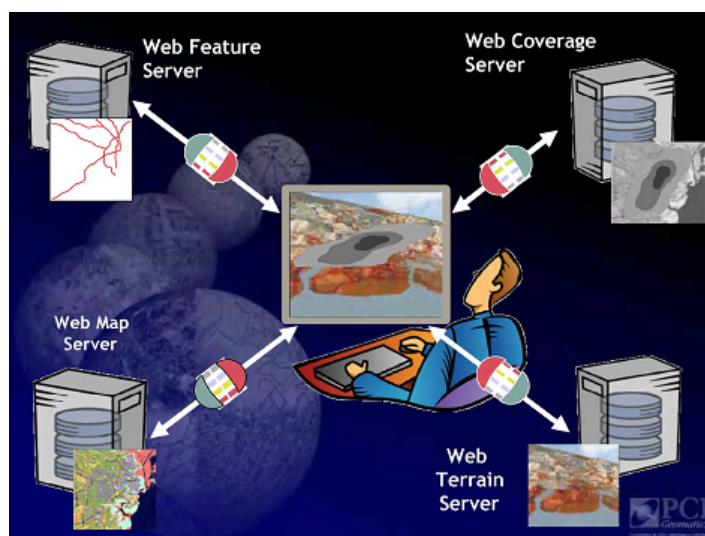


Figura 2.5: Serviços de acesso a mapas na Internet. Fonte: *PCI Geomatics* [75].

No caso dos *WMS*, para que os dados sejam compartilhados com independência da aplicação, um conjunto básico de “metadados” deve ser definido conforme a Tabela 2.2. Quando se trata de outros serviços, como o *WFS* e o *WCS*, outros “metadados” são definidos, os quais permitem à aplicação conhecer os recursos oferecidos pelos serviços, bem como lhe capacita a exercer controles sobre os aspectos temporais, espaciais e dos atributos. Assim, é possível requisitar um subconjunto importante para a visualização geográfica que se deseja

no momento.

Nome	Descrição	Exemplo em WMS
Tipo de serviço	Indica que a requisição será para o WMS	service=WMS
Requisição de mapa	Indica que a requisição será de um mapa	request=getmap
Versão do WMS	Especifica a versão que deve ser compatível com a do cliente	version=1.1.1
Projeção ou sistema de referência espacial	Escolhe o sistema de projeção da imagem georreferenciada	srs=EPSG:4326
Formato de Imagem	Escolhe o formato de imagem a ser fornecido pelo servidor	format=image/jpeg
Nomes dos Layers ou fontes de dados	Seleciona as camadas que serão enviadas	layers=Estado,Municípios
Tamanho da imagem	Especifica o tamanho da imagem em largura e altura, em pixels	width="1000" height="740"
Extensão geográfica ou <i>bounding box</i>	Expressa as coordenadas sudoeste e nordeste do mapa	bbox=-170 0,-50 90

Tabela 2.2: Tabela com os principais metadados usados no padrão WMS.

## 2.8.1 Protocolos de Comunicação dos Serviços Web

Os dados obtidos dos serviços *Web* são transmitidos usando-se protocolos de comunicação recomendados pelo OGC. Tais protocolos apresentam um conjunto de procedimentos cujos objetivos são: estabelecer uma forma comum de representação de dados, mensagens e descrições de serviços; permitir a extensão de mensagens e serviços; e criar um mecanismo para a localização de serviços. Alguns procedimentos pertinentes a este trabalho são descritos nas próximas subseções.

### 2.8.1.1 Protocolo SOAP

O protocolo SOAP, do inglês, *Simple Object Access Protocol*, é um protocolo para troca de informações em um ambiente descentralizado e distribuído. As trocas de mensagens entre diferentes serviços são padronizadas através da definição do formato das mensagens XML.

Parte da sua especificação é composta por um conjunto de regras de como usar o documento XML para representar os dados. Outra parte define o formato de mensagens, convenções para representar as chamadas de procedimento remoto (RPC) e associações ao protocolo HTTP. O protocolo contém os seguintes elementos:

- *Envelope*: identifica o documento XML como uma mensagem SOAP e é responsável por definir o conteúdo da mensagem.
- *Header* (opcional): contém os dados do cabeçalho.
- *Body*: contém as informações de chamada e de resposta ao servidor.
- *Fault* (somente nas mensagens de resposta do servidor): contém as informações dos erros ocorridos no envio da mensagem.

### 2.8.1.2 Gramática WSDL

A gramática WSDL, escrita na linguagem XML, define serviços como uma coleção de *endpoints* ou portas, provendo uma padronização para a descrição da funcionalidade desses serviços, definindo o conteúdo das mensagens de requisição e das mensagens de resposta, isto é, declarando quais mensagens o serviço *Web* aceita e gera. Tudo isto estabelece um contrato do serviço *Web*. Neste contrato são definidos, além das entradas e saídas do serviço, os esquemas XML que podem ser usados para a criação dos documentos de entrada e saída. A WSDL mostra o caminho para que clientes ou outros serviços acessem eletronicamente o serviço definido. Os elementos *message*, *operation* e *portType* provêm uma definição abstrata para os dados que estão sendo trocados e as operações que estão sendo executadas pelo serviço. Cada ligação define como mapear os tipos de portas a um conjunto de portas concreto, usualmente pela provisão de codificação e parâmetros de conexão específicos a uma instância do serviço. A WSDL é constituída pelos seguintes elementos:

- Tipos - um *container* para definições de tipos de dados que usa algum sistema de tipo, como, principalmente, o XML Schema XSD [110].
- Mensagem (*message*) - uma definição abstrata e com os tipos dos dados transmitidos.

- Operação (*operation*) - uma descrição abstrata de uma ação suportada pelo serviço.
- Tipo de porta (*portType*) ou interface - um conjunto de operações suportadas por uma ou mais portas.
- Ligações (*bindings*) - uma especificação concreta de protocolo e formato de dados para uma interface particular. Por exemplo, a definição de que o serviço pode ter seu acesso via algum dos protocolos HTTP Get, HTTP Post ou SOAP, sendo o mais usado o SOAP.
- Porta - um simples *endpoint* definido como uma combinação de uma ligação e um endereço de rede.
- Serviço - uma coleção de portas associadas.

### 2.8.1.3 Outros Protocolos dos Serviços Web

Um conjunto de aplicações SOAP denominado UDDI, do inglês *Universal Description, Discovery, and Integration* [66] que permite a publicação e a localização de serviços *Web*. Através deste mecanismo, os servidores avisam sobre a existência de seus serviços e os clientes localizam os serviços de seu interesse.

O protocolo RPC, do inglês *Remote Procedure Call*, que pode ser usado por uma aplicação para requisitar um serviço de outra aplicação localizada em outro computador na rede, sem que haja a necessidade de reconhecimento de detalhes da rede.

A abordagem de documento literal, também conhecido como mensagem de chamada, que consiste no envio de um documento XML como uma mensagem de requisição e na resposta retornada pelo servidor em outro documento XML.

Pode-se definir, de forma simplificada, um serviço *Web XML* como um serviço publicado através do SOAP, descrito com um arquivo WSDL e registrado em UDDI [57].



## 2.9 Mapas Colaborativos na Internet

Uma iniciativa emergente do uso desta tecnologia é a produção em conjunto e *online* de modelos de localizações do mundo real. A idéia, derivada das aplicações para publicações pessoais (*blogs*) e para edição de documentos em grupo (*wikis*), baseia-se na ação coletiva de usuários que podem, além de visualizar, inserir representações de objetos georreferenciados. Chang [16] visiona a *Web* como um meio de lugares, pessoas e participação. Há várias razões para o sucesso da *Web* como um meio para a colaboração aberta. Ela provê acesso a uma coleção rica de recursos *online* os quais agem como uma base comum de colaboração.

A pesquisa participativa é uma abordagem da Geografia Social que oferece um modelo para o desenvolvimento e envolvimento comunitário. Esta pesquisa não somente indica que a produção de mapas é uma das melhores técnicas de participação como também reconhece que esta abordagem favorece:

- a construção de comunidades ou redes de relacionamentos com a reunião de pesquisadores do mundo todo;
- uma abordagem colaborativa e não hierárquica para a investigação de inter-relacionamentos nos domínios do conhecimento;
- a integração de idéias e contribuições de diferentes atores em projetos comunitários; e
- um meio de participação para qualquer integrante da comunidade.

## 2.10 Proveniência e Fluxo dos Dados

Quando vários profissionais manipulam os dados em um trabalho colaborativo é desejável que seja possível reproduzir os procedimentos em que os dados originais foram transformados ou combinados para a criação de visões da base de dados. A proveniência dos dados, que permite descobrir de onde veio um conjunto de dados e o processo pelo qual este chegou na base de dados, tem se tornado importante, especialmente em bases de dados científicas onde a compreensão da linhagem dos dados é crucial para comprovar a sua acuracidade e promover a aceitação dos mesmos. Somente uma parte dos dados existentes em uma base de dados pode

ser considerada como fonte no sentido em que foi obtida a partir de experimentos. Os outros dados são visões, ou seja, foram obtidos a partir de outras fontes de dados ou de outras visões. O entendimento de que as visões não são produzidas unicamente por meio de consultas, mas também de correções e anotações inseridas por especialistas, aumenta a complexidade deste controle [77].

Bowers e outros [96] propõem um modelo para o gerenciamento da proveniência e a análise dos dados em fluxos de trabalhos científicos, onde as operações de leitura (*Read*), escrita (*Write*) e o estado de reinício (*State-reset*) (*RWS*) de um processo são gravados em uma tabela relacional, sobre a qual são programadas consultas capazes de reconstruir completamente os dados e de obter grafos de dependências entre os mesmos. No caso mais simples, o fim de uma rodada ou de uma unidade lógica de trabalho é marcado pela mudança da ação de escrita para a ação de leitura, mas há casos em que esta mudança não é suficiente para marcar o estado de reinício. Isto ocorre quando há dependências de resultados anteriores para os cálculos que envolvem séries históricas ou onde são executados filtros sobre as leituras e escritas precedentes. Nesses casos, o mecanismo para a detecção do início de uma nova rodada pode ser obtido do próprio modelo de computação aplicado. Os autores adotam as noções e terminologia do sistema de fluxo de trabalho científico Kepler [47], onde fluxos de trabalhos são compostos por “atores” conectados em um grafo. A comunicação entre os “atores” é feita por meio de portas, onde portas de saída podem ser conectadas a portas de entrada, estabelecendo canais unidirecionais. Os “atores” se comunicam por estes canais passando “tokens”, os quais são imutáveis e duram somente entre a sua criação na porta de saída e o seu consumo na porta de entrada. Mesmo que um “ator” somente repasse os dados, sem que os mesmos tenham sido alterados, uma nova identificação para o *token* é criada com o objetivo de facilitar o rastreamento da dependência entre os mesmos. Este *log* de eventos para um fluxo de trabalho é usado para a obtenção de gráficos de dependência dos *tokens*, por meio dos quais se pode saber quais os que contribuíram para a produção de um determinado *token*. O modelo de computação de um fluxo de trabalho é especificado por um componente chamado “diretor”, o qual especifica e media todas as comunicações inter-atores. Exemplos de “diretores” são as semânticas de rede de processo, de evento discreto, de tempo contínuo

e de transdutores de estado finito.

## 2.11 Sistemas Georreferenciados Aplicados à Saúde

O primeiro emprego de mapas para a descoberta ou confirmação da causa de uma epidemia que se tem notícia foi publicado pelo Dr. John Snow [12]. Em 1849, o Dr. Snow publicou um panfleto onde propunha que o cólera reproduzido no corpo humano era espalhado por alimentos ou água contaminada. Esta teoria sofreu oposição da mais comumente idéia aceita na época de que o cólera, como todas as doenças, era transmitido através da inalação de vapor contaminado. Em 1854, quando o cólera atingiu a Inglaterra novamente, o Dr. Snow conseguiu legitimar seu argumento. A investigação epidêmica iniciou com a plotagem das mortes relacionadas com o cólera. Naquela época, a água de Londres era suprida por duas companhias. Uma bombeava a água da superfície do rio Tâmesa, enquanto a outra bombeava a água subterrânea de poços. Uma alta concentração de cólera foi encontrada na região da cidade que usava a água subterrânea. Em seguida, ele constatou que, em uma determinada localização, perto da intersecção entre as ruas Cambridge e Broad, havia ocorrido mais de 500 mortes por cólera em 10 dias. Este foi um dos primeiros usos de múltiplos mapas temáticos, técnica que consiste na sobreposição gráfica de camadas de dados em uma série de mapas de base, o que possibilita o relacionamento de objetos e fenômenos geograficamente. Neste mapa apresentado na figura 2.6, com a visualização de barras empilhadas nos locais de residência e proporcionais à quantidade de óbitos causados pelo cólera de moradores destes endereços, obteve-se a indicação do local da fonte contaminante.

O uso de recursos computacionais para o mapeamento temático e a análise espacial iniciou-se na década de 1960, quando o *Canadian Geographic Information System (CGIS)* [15] e o *Harvard Laboratory for Computer Graphics and Spatial Analysis* [17] desenvolveram os primeiros sistemas aplicados a esta finalidade. Até os anos 1980, o uso dos SIG era restrito às universidades ou ao governo, onde o software era executado em *mainframes*. No ano de 1982, a *Environmental System Research Institute, Inc (ESRI)* [32] lançou o seu primeiro software SIG comercial, chamado ARC/INFO que era executado em minicomputadores. Com o aumento da capacidade do hardware e a redução de seu preço foi possível, nos anos 1990,



Figura 2.6: Mapa empregado na pesquisa epidêmica do cólera na cidade de Londres em 1854 - Fonte: CSISS [12]

o desenvolvimento de SIG para microcomputadores, quando esta tecnologia passou a ter uma maior usabilidade provida pelos recursos gráficos que surgiram neste ambiente computacional. Isto também tornou a tecnologia acessível a um número maior de profissionais que instalaram a primeira versão comercial de SIG para *desktop*, o ArcView [6]. No primeiro semestre de 1992, a ESRI vendeu 10 mil cópias do ArcView. O próximo passo para o aumento da acessibilidade aos mapas digitais foi o lançamento, no ano de 1999, do ArcIMS, o único software SIG da época que permitia aos usuários integrar dados locais com dados de Internet em uma interface simples de um navegador (*browser*).

Nos últimos anos, com a popularização da *World Wide Web* ou *Web*, a academia e a indústria voltaram-se para o desenvolvimento dos chamados sistemas de localização para *Web*, onde mapas digitais são acessados por clientes conectados à rede com o uso de um simples navegador e os dados espaciais são fornecidos por servidores de mapas. Esta tecnologia, chamada de *Web Mapping* [58], inicialmente foi baseada no modelo de interface tradicional para uma aplicação de Internet, em que o usuário requer uma página ao servidor, o qual a constrói e a entrega ao navegador. Esta página inicial inclui algum formulário HTML usado para a entrada de dados pelo usuário. Uma vez que o formulário preenchido é enviado, a próxima página é construída e servida de acordo com as opções selecionadas pelo usuário. Este processo continua em um estilo de navegação ditado pela natureza do protocolo HTTP. A principal desvantagem deste modelo em relação ao modelo de aplicação dos *desktops* é a espera pela atualização das páginas pelo usuário, o que faz com que haja uma interrupção no fluxo de pensamento relacionado a sua ação. Nos *desktops*, isto não acontece porque a interface com o usuário está inerentemente conectada à camada de aplicação.

Com o incremento da velocidade de transmissão na Internet, um conjunto de tecnologias passou a ser empregado para que o problema da quebra do fluxo da ação do usuário fosse evitado. Este conjunto de tecnologias foi batizado por Jesse James Garrett, da Adaptive Path, como AJAX, do inglês *Asynchronous Javascript and XML* [34]. Esta tecnologia usa um recurso implementado nos navegadores modernos conhecido como objeto XMLHttpRequest, o qual habilita o código JavaScript a fazer requisições a um servidor remoto, sem a necessidade de recarregamento da página, evitando as interrupções e aumentando a eficiência do sistema. Diante de uma interface eficiente, o usuário concentra-se na lógica de suas ações, ocupando-se exclusivamente com suas tarefas, sem desviar a atenção para o que está sendo feito pela camada da aplicação. Este trabalho emprega um *framework* que implementa a tecnologia AJAX.

Um exemplo de visualização geográfica na *Web* é encontrado no *Atlas of Cancer Mortality in the United States* [63, 27]. Neste atlas pode-se notar a distribuição espacial da taxa de mortalidade por câncer, considerando-se todos os cânceres ocorridos nos Estados Unidos, entre 1970 e 1994, discriminados por sexo e raça. Por exemplo, nas figuras 2.7 e 2.8, que

representam a taxa de mortalidade por câncer, respectivamente, em negros e brancos, pode-se verificar que entre os homens brancos há uma maior taxa de mortalidade por câncer na região sudeste, enquanto que entre os homens negros há uma maior taxa de mortalidade por câncer na região central dos Estados Unidos.

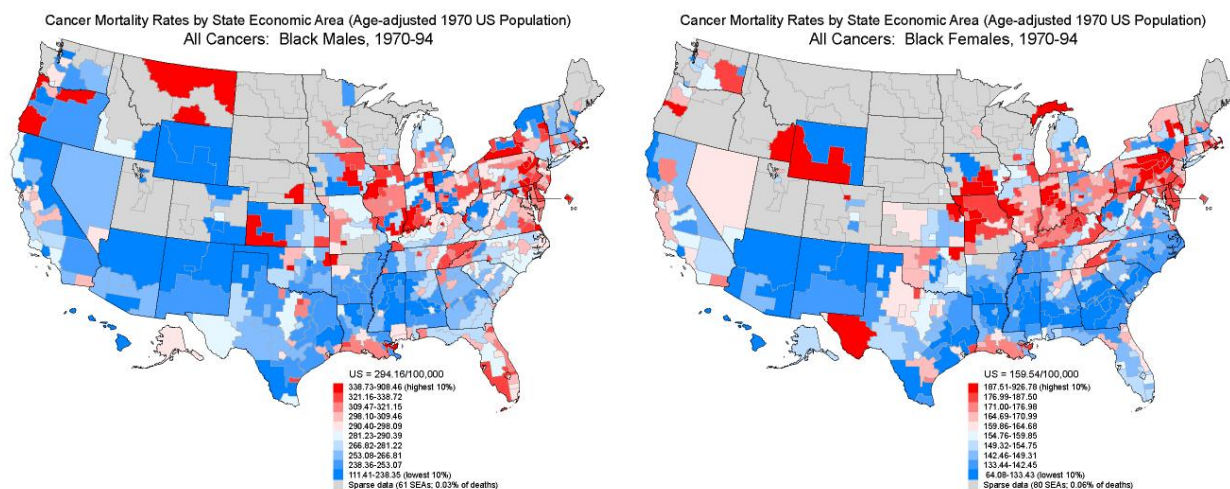


Figura 2.7: Mapa da mortalidade por câncer em negros, homens e mulheres, nos Estados Unidos, entre 1970 e 1994. Fonte: Atlas of Cancer Mortality in the United States [27, 63].

Sabe-se que a densidade das populações negra e branca se distribui de maneira heterogênea, porém é possível que uma exploração das figuras 2.7 e 2.8 já possibilitaria a um pesquisador da epidemiologia formular inúmeras hipóteses sobre as associações da genética (raça) com fatores do ambiente conhecidos como, por exemplo: o clima, os hábitos culturais dos habitantes (como a alimentação) e o uso do solo na agricultura, nas regiões onde a incidência de mortes por cânceres foi mais evidente. Outras hipóteses poderiam enriquecer a abrangência de uma pesquisa caso este pesquisador dispusesse de outras variáveis do ambiente, conhecidas somente em ciências como a Geologia e a Química, como é o caso dos elementos químicos biodisponíveis. Uma ciência emergente, a Geologia Médica [93], a qual examina os impactos de materiais e processos geológicos na saúde humana e do ecossistema, incluindo ambas as fontes de potenciais problemas à saúde, naturais e antropogênicas, é a principal beneficiada com este projeto.

No Brasil têm-se o conhecimento de alguns trabalhos relacionados à Geologia Médica e à Epidemiologia, como o projeto Paisagens Geoquímicas e Ambientais do Vale do Ribeira [37], desenvolvido através de uma parceria entre o Departamento de Geociências e a Faculdade

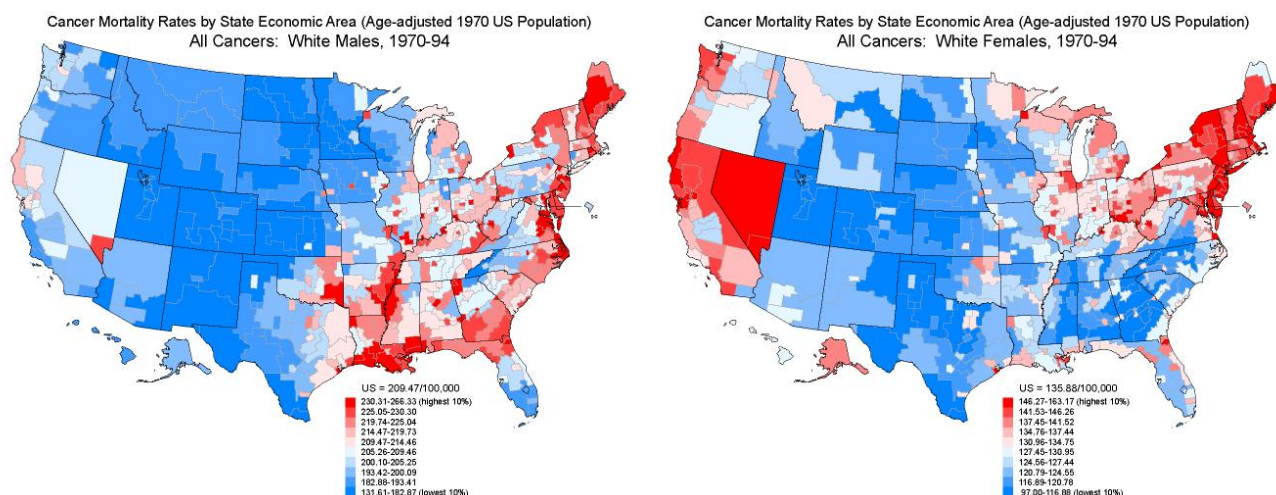


Figura 2.8: Mapa da mortalidade por câncer em brancos, homens e mulheres, nos Estados Unidos, entre 1970 e 1994. Fonte: Atlas of Cancer Mortality in the United States [27, 63].

de Ciências Médicas da Universidade Estadual de Campinas (UNICAMP), a Universidade Estadual de Londrina (UEL) e o Instituto Adolfo Lutz; e o projeto SAUDAVEL [91] (Sistema de Apoio Unificado para a Detecção e Acompanhamento em Vigilância Epidemiológica), uma parceria entre a Divisão de Processamento de Imagens do Instituto Nacional de Pesquisas Espaciais (INPE), o Laboratório de Estatística Espacial da Universidade Federal de Minas Gerais (UFMG), o Departamento de Estatística da Universidade Federal do Paraná (UFPR); a Fundação Oswaldo Cruz, o Departamento de Epidemiologia e Métodos Quantitativos em Saúde da Escola Nacional de Saúde Pública e o Departamento de Saúde Coletiva do Centro de Pesquisas Aggeu Magalhães.

O projeto “Paisagens Geoquímicas e Ambientais do Vale do Ribeira” [60] avaliou o risco de contaminação por elementos químicos naquela região. Descobriu-se que havia um nível elevado do elemento chumbo no solo do município de Adrianópolis, Paraná, e que a fonte poluidora foi a atividade de beneficiamento e refino de chumbo praticadas entre 1945 e 1995. Uma vez verificada a contaminação de solos por metal, uma segunda pesquisa consistiu em analisar os alimentos semeados nesses solos e consumidos pelos moradores. Além da contaminação dos vegetais, constatou-se também que a população que residia nas proximidades da mineradora desativada, ainda era contaminada pelo chumbo presente na superfície do solo, por meio da poeira, apresentando concentrações no sangue superiores aos níveis internacionalmente aceitos. Esta pesquisa recomendou aos responsáveis pela saúde pública a necessidade de

cuidados especiais com a saúde da população afetada e de educação para evitar a exposição dos indivíduos nas áreas de risco.

As questões relacionadas à eficácia da comunicação do risco do projeto “Paisagens Geoquímicas e Ambientais do Vale do Ribeira” foram estudadas por Giulio [41], que avaliou o impacto sofrido pela comunidade, posteriormente às informações veiculadas pela imprensa. A comunicação de risco foca a troca de informações entre diferentes atores sociais, basicamente no que se refere à natureza, magnitude, interpretação e gestão do risco. Nesta comunicação é preciso considerar todas as facetas, uma vez que a própria pesquisa do risco já é de natureza alarmante. Através de pesquisa da literatura, empírica e documental, entrevistas com pesquisadores, jornalistas, moradores e autoridades do município, Giulio concluiu que esta comunicação pode aumentar crenças e mitos e reforçar rótulos criados em torno de pessoas e locais.

No caso de um município onde foi comprovada a contaminação, os moradores foram rotulados de “chumbosos” e tiveram dificuldades quanto à admissão em empregos em outros locais. Também houve desvalorização de suas propriedades, sendo que inicialmente os produtos agrícolas foram rejeitados por consumidores de outros municípios. Giulio recomenda uma estratégia em que a população dos lugares contaminados participe ativamente, desde o início da pesquisa, do processo decisório para a solução dessas situações consideradas de risco.

Uma experiência piloto do projeto SAUDAVEL [91] é aplicada no sistema de vigilância, controle e alarme do risco da dengue na região metropolitana do Recife, Pernambuco, onde, através do registro semanal da quantidade de ovos do mosquito *Aedes spp* coletados em pontos intra-urbanos, avalia-se o risco desta epidemia com o emprego intensivo de tecnologias da informação integradas para o tratamento de informações espaço-temporais e modelos preditivos. A proposta do SAUDAVEL é de modificar a natureza dos sistemas de vigilância em saúde no Brasil, os quais apresentam uma natureza reativa e não preemptiva, considerando que geralmente a detecção e o acompanhamento são em função da incidência de casos da doença. Assim, as tecnologias da informação possibilitam a incorporação de aspectos ambientais identificadores de riscos e produzem os instrumentos tecnológicos necessários à antecipação e, conseqüentemente, à ampliação da capacidade preventiva.



Nestas duas experiências brasileiras foram usadas, para a visualização geográfica, ferramentas que não disponibilizam as informações em um ambiente de *Web Mapping*, mas em um ambiente de *desktop*, como o Spring [20], o TerraView [105], e o ArcView<sup>TM</sup> que, além da visualização dos dados georreferenciados, permitem também a interatividade do usuário, mas exige um treinamento para o uso e também a instalação das ferramentas em seu *desktop*. A grande vantagem de uma ferramenta com a tecnologia *Web Mapping* é que esta não requer a instalação de software nos computadores dos usuários, além de ser constituída por uma interface amigável, a qual o usuário está familiarizado, como é o caso da interface *Web*, facilitando o aprendizado para o uso da ferramenta de visualização geográfica.

Vários trabalhos sobre sistemas georreferenciados aplicados à saúde foram publicados no *International Journal Health Geographics (IJHG)* [44] nos últimos anos, dos quais, nos trabalhos de Boscoe e outros [11] e Bell e outros [8] verificam-se características importantes para o GeoMedicina, apresentadas nos próximos parágrafos.

No trabalho de Boscoe e outros são analisadas as diversas fontes de dados empregadas em estudos geográficos do câncer dos Estados Unidos, como registros de notificação de câncer, censos demográficos, serviços de vigilância, meio ambiente, sensoriamento remoto e fotogrametria aérea. Ao discorrer sobre a qualidade dos dados espaciais, os autores alertam sobre particularidades destes dados e tratamentos apropriados, sendo os mais relevantes a este trabalho listados a seguir:

- Além do espaço canônico, devem ser considerados os espaços relativos, nos quais a distância é definida em termos de outro atributo, como similaridade demográfico-social ou conectividade ao longo de redes de transporte. O destino e transporte de químicos no ambiente devem ser considerados. Uma simples proximidade não indica que um indivíduo desenvolveu certa doença por estar exposto a um determinado elemento contaminante.
- Outros fatores, além do espacial, devem ser verificados, como a comprovação biológica de associações entre a exposição e a doença, a diferente qualidade da água no ambiente e na rede de abastecimento, sendo esta mais consumida por populações urbanas e a primeira mais consumida por populações rurais.

- O período de latência longo para muitos cânceres e as migrações da população americana sugerem que endereços anteriores sejam considerados e não somente o endereço atual.
- Ao considerar somente o endereço de residência do indivíduo na análise, ignora-se fatores de risco relacionados ao seu comportamento e hereditariedade, como alimentação, tabagismo, locais de lazer e trabalho. Características como cultura, profissão, nível social, idade, sexo, raça, etnia devem ser consideradas para que a análise seja completa.
- Para concluir que uma rua ou um bairro de uma cidade apresenta uma incidência alta de câncer para uma determinada classe (idade, sexo, raça) de habitantes, é necessário considerar a quantidade de habitantes com as mesmas características que não desenvolveram este câncer na mesma região.
- O problema de inferência com dados espacialmente desalinhados ou problema de sobreposição de polígonos deve ser solucionado com o emprego de métodos como o de informação covariante ou inferência Bayesiana hierárquica.
- Outra característica importante é a da não estacionaridade espacial, que é a tendência de variação dos relacionamentos por localização geográfica. As estatísticas globais ignoram esta característica. Para considerá-la, devem ser empregadas as estatísticas locais.
- A instabilidade da variância, que é a característica das taxas de doenças em áreas com poucos casos serem mais variáveis que aquelas em áreas com muitos casos, causa um erro sistemático de visualização em um mapa coroplético, como, por exemplo, a tendência da visualização de áreas rurais, onde a população é esparsa, dominar a visualização de áreas urbanas densamente povoadas.

No trabalho de Bell e outros [8] são abordadas técnicas para o *design* de mapas que garantam uma comunicação eficiente de estatísticas sobre o câncer na população. A natureza desta comunicação envolve uma série de peculiaridades que devem ser consideradas para que a comunicação seja eficiente. O maior desafio desta atividade é assegurar que os mapas de estatísticas sobre a saúde informem e não confundam. A seguir são apresentados alguns aspectos importantes para que esta tarefa seja executada com êxito:

- O propósito dos mapas sobre doenças está associado a questões acerca da taxa de incidência em certa área, da identificação de alguma tendência que denote um padrão ou da similaridade entre padrões. Recomenda-se que não se tente responder a muitas questões em um único mapa, pois mapas que têm mais de um propósito ofuscam a mensagem e confundem o leitor.
- Ao mesmo tempo em que se busca a simplicidade na representação em mapas, para se aumentar a confiabilidade da informação, principalmente para uma audiência científica, é necessário expressar detalhes da metodologia, vantagens e desvantagens de uma análise específica e informações sobre: fonte dos dados, título, legenda, escala, etiquetas para nomes de locais ou valores de atributos, sistema de projeção empregado, data da produção. Estas informações sobre a qualidade dos dados podem ser incluídas em metadados definidos em padrões de interoperação.
- Um cuidado especial deve ser tomado quanto às inferências baseadas em comparações de dados agrupados. Associações observadas em um nível de agregação podem ser inconsistentes com as observadas em um nível de indivíduos. Por exemplo, mesmo que se infira que as taxas de mortalidade por câncer de pulmão e de fumantes são elevadas em uma determinada região, não se pode saber a priori se aqueles indivíduos em que a causa de óbito foi câncer de pulmão eram fumantes.

## 2.12 Sistemas de Web Mapping para Computação Geoespacial

O *Web Mapping* consiste no desenvolvimento de aplicações para ambiente *Web* que propiciam ao usuário a capacidade de visualizar e interagir com a informação geográfica. Assim, é possível aliar a facilidade de uso de uma interface *Web* à facilidade de interpretação da representação de dados visuais (mapas, gráficos, relatórios formatados) em uma solução simples para publicação de dados para o usuário final [51].

A disseminação de informações geográficas pela Internet aumenta a capacidade de acesso aos mapas. A manutenção distribuída das bases de dados espaciais e a interoperação entre os sistemas criam um ambiente propício para o trabalho colaborativo, onde a informação

atualizada por uma fonte é servida a multiusuários, em uma entrega rápida e em uma atualização facilitada da informação.

Uma arquitetura em camadas fornece uma visão do software usado em um sistema para computação geoespacial em plataforma *Web*, desde o software de base até a aplicação. A figura 2.9 apresenta, lado a lado, a arquitetura genérica e a arquitetura dos softwares de código livre e aberto para computação geoespacial ou, em inglês, *Free Open Source Software for Geoinformatics* (FOSS4G), agrupados nas camadas sistema, processamento de dados, serviço de dados, interface com o usuário e aplicação do usuário final [46].

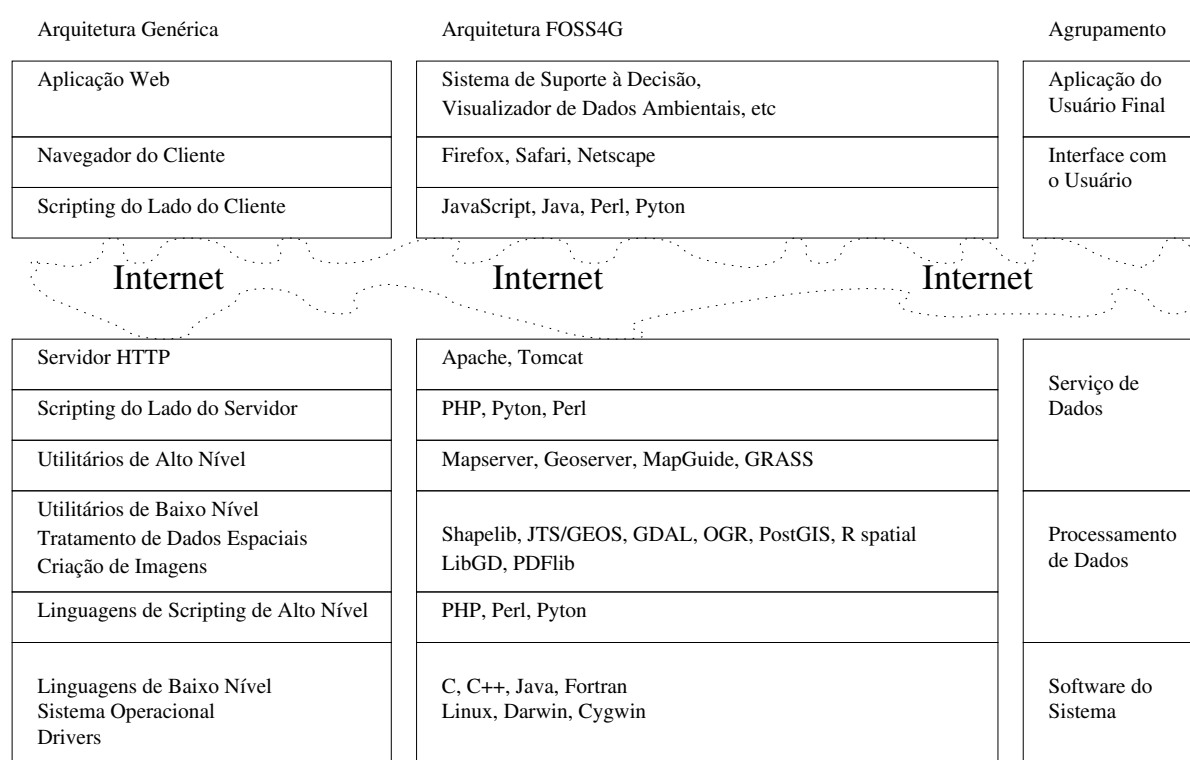


Figura 2.9: Arquitetura em camadas para FOSS4G em plataforma *Web*.

A camada sistema agrupa o sistema operacional, cuja opção natural para o FOSS4G é o Linux embora o *framework* da Microsoft .NET tenha criado uma oportunidade interessante para o desenvolvimento de software de código aberto e livre. As linguagens C, C++ e Java são usadas para o desenvolvimento de bibliotecas e ferramentas de baixo nível, porém a opção pelo Java dificulta a integração com outras linguagens, o que faz com que esta escolha leve a uma solução com o uso predominante desta linguagem nas camadas superiores da plataforma do servidor [46].

A camada de processamento de dados engloba o gerenciamento, o processamento de formato e análise dos dados espaciais. Esta camada é responsável pela leitura, interpretação e transformação dos objetos geográficos organizados em diversos formatos. Um exemplo de software desta camada é a biblioteca GDAL, do inglês *Geospatial Data Abstraction Library* [36] a qual pode ser usada para acessar, ler, manipular e escrever em mais de 50 formatos *raster*. A *OGR Simple Features Library* [73] é uma parte da GDAL que faz o mesmo para mais de 20 formatos vetoriais. Outro exemplo de software usado para este processamento é a ferramenta que transforma conjuntos de dados geográficos de um sistema de coordenadas para outro. No FOSS4G a biblioteca mais comumente usada para esta finalidade é o PROJ4 - *Cartographic Projections Library* [86].

A análise é realizada com ferramentas de computação estatística ou geométrica, como é o caso da linguagem R de computação estatística e a *Geometry Engine Open Source* (GEOS) [39] a qual provê operações como sobreposição espacial e predicado binário. Os SGBD espaciais são usados para a organização de dados vetoriais, com o aproveitamento das propriedades destes sistemas e a integração de bibliotecas como a GEOS e a PROJ4.

Ainda na camada de processamento, há bibliotecas usadas para produzir as imagens resultantes da interpretação e da análise dos dados espaciais. Vários formatos de imagens são produzidos como as imagens JPG, GIF, PNG e PDF. A biblioteca *GD Graphics Library* (GD) [35], usada para a criação dinâmica de imagens e a biblioteca *PDFlib* [76], usada para a criação do formato PDF/A especificado no padrão ISO 19005, são exemplos do software usado na produção destes formatos de imagens.

A camada de serviço engloba o servidor de mapas e o servidor HTTP. O servidor de mapas organiza as camadas de mapas obtidas com o processamento das bibliotecas da camada anterior repassando-as ao servidor HTTP. Os servidores de mapas integram diferentes fontes de dados espaciais através das bibliotecas que as lêem e da adequação a padrões de interoperação. Exemplos de servidores de mapas são o MapServer [108], o GeoServer [40] e o MapGuide [52]. O servidor HTTP recebe as requisições da camada interface com o usuário, ativa a execução do servidor de mapas e de interpretadores de *scripts* e serve os resultados empacotando-os no protocolo HTTP. Estas ferramentas servem mapas como imagens, o que viabiliza a criação de

*Websites* interativos.

A camada interface com o usuário consiste na interface dos *browsers* usados para navegação na Internet, a qual oferece controles de navegação, processamento de tipos de mídia e processamento de código, como JavaScript, XML e GML. Um tipo de mídia, do inglês *Media Type* ou, segundo o termo original, *Multipurpose Internet Mail Extensions* (MIME) [55], antecede os diferentes conteúdos recebidos pelo *browser* e serve para acionar o processamento apropriado a estes dados. Por exemplo, um texto em código HTML é indicado com MIME “text/html”, já uma imagem GIF é indicada com “image/gif”.

A camada de aplicação do usuário final é obtida com a execução no *browser* do código recebido do servidor HTTP. Ela é constituída por controles da funcionalidade do sistema e pela visualização dos resultados obtidos com a interação do usuário.

O próximo capítulo aborda as características de componentes FOSS4G que constituem o protótipo inicial do GeoMedicina, bem como um projeto de evolução deste protótipo com a aplicação dos estudos constantes nesta revisão bibliográfica.

## CAPÍTULO 3

### PROJETO E EVOLUÇÃO DO SISTEMA

No capítulo 2 foram revistos os tópicos de Engenharia de Software, Cartografia e Bases de Dados que contribuem para a qualidade de um sistema *Web mapping*, onde se abordou o emprego de software livre com a integração de componentes que constituem um protótipo inicial. No modelo espiral viu-se como o protótipo inicial evolui para atender às necessidades dos usuários do sistema. Observou-se também a importância de seguir padrões internacionais para a garantia da expansão do sistema.

Neste capítulo é apresentado o levantamento de requisitos do sistema, os quais levaram à escolha dos componentes do protótipo inicial, que é uma interface de programação de aplicativos (do inglês *Application Programming Interface - API*) denominada Ka-map [28]. O projeto de evolução e manutenção deste protótipo com a agregação de novas funcionalidades é apresentado em uma rede de atividades. Nestas atividades estão previstas a elaboração de um diagrama entidade-relacionamento, o aperfeiçoamento da interface com o usuário e a implementação de algoritmos nas linguagens PHP e SQL para prover a funcionalidade esperada do sistema.

Para o sucesso do projeto é fundamental que o sistema atenda às expectativas dos usuários, as quais são identificadas mediante um levantamento e análise de requisitos. Nesta análise se procura determinar o máximo de informações sobre o domínio da aplicação e organizar os requisitos levantados, classificando-os em grupos coerentes, dentro de uma prioridade e isentos de conflitos. No levantamento de requisitos, a equipe de desenvolvimento absorve vários pontos de vista de usuários heterogêneos através de revisão bibliográfica dos sistemas correlatos, de reuniões com a direção do projeto, de participação em simpósios ou *workshops*, ou através de entrevistas com os usuários potenciais do sistema. Estes pontos de vista são organizados de forma a mostrar informações detalhadas sobre os serviços oferecidos pelo sistema, os dados que o sistema requer e como ambos são controlados.

Neste processo de software é necessário conciliar requisitos de usuários heterogêneos com requisitos do sistema, derivados dos modelos de domínio e da API empregada no protótipo inicial.

### **3.1 Escopo do Sistema**

O GeoMedicina é um sistema multiusuário e distribuído via rede de computadores, desenvolvido com software livre e de código aberto, cuja versão inicial é dimensionada para acesso simultâneo de até cem usuários. O tempo máximo de resposta permitido para cada requisição de usuário é de um minuto. O sistema apresenta uma visualização geográfica de elementos químicos e das doenças na região de estudo e a sobreposição destes temas.

A apresentação dos mapas digitais é feita através de imagens em páginas de hipertexto, as quais apresentam uma interface com o usuário com controles de navegação e de seleção de aspectos de interesse. Para isso, é instalado um servidor de mapas de alta disponibilidade.

No projeto GeoMedicina, o reuso de componentes minimiza as atividades do processo de software, uma vez que a especificação, o projeto e a implementação da primeira versão do software já estão incorporadas em uma API que constitui o primeiro protótipo. Neste trabalho é necessária uma estruturação da base de dados espacial, a configuração e a instalação de um protótipo e a evolução deste protótipo com a integração de novos subsistemas e com o aperfeiçoamento das interfaces, em refinamentos com repetidas interações entre os usuários e os projetistas. A iteração do processo de software neste trabalho segue o modelo de desenvolvimento incremental.

### **3.2 Cuidados com os Riscos do Projeto**

Um gerenciamento seguro de projeto requer que sejam identificados os riscos, a probabilidade de ocorrência e o impacto causado por essa ocorrência. Caso contrário, o projeto estaria susceptível a sofrer com os empecilhos decorrentes de problemas irrevogáveis que poderiam até causar o seu fracasso. A seguir são colocados em ordem de prioridade, do mais importante para o menos importante, os riscos analisados no projeto GeoMedicina, bem como as medidas



de administração dos riscos tomadas para a reversão dos mesmos:

- Atraso na entrega da primeira versão, devido a complicações em integrações de componentes ou em adaptações da interface com o usuário do protótipo. Para evitar este risco é estabelecida uma rede de tarefas associada a um cronograma no qual constam as tarefas e os marcos de referência para o controle do gerente de software.
- Pouca confiança dos usuários nos resultados apresentados, devido à inconsistência da interface ou à ausência de validação dos valores quantitativos. Isto é evitado com o emprego de técnicas da engenharia semiótica na construção da interface, a qual apresenta etiquetas com valores numéricos posicionadas sobre os respectivos objetos geométricos do mapa digital.
- Desempenho baixo, relativo ao tempo de respostas às requisições do cliente, devido a algum gargalo no hardware, rede ou algoritmo empregado em alguma camada da arquitetura. O protótipo inicial tem um desempenho superior ao especificado no escopo de sistema. Em cada integração de componente de software deve ser feito um monitoramento do desempenho para que a adição de novas camadas não eleve o tempo de espera para valores inaceitáveis.
- Comprometimento da integridade do sistema, devido à falta de segurança ou falhas de proteção. O emprego de um controle de acesso com senha e a observação de itens referentes à proteção e à segurança na configuração do servidor de mapas são cuidados necessários para evitar qualquer vulnerabilidade do sistema.

### **3.3 Requisitos da Base de Dados**

A modelagem da estrutura de dados do GeoMedicina considera a sua natureza espacial, organizando o acesso e as relações entre os dados de forma a garantir propriedades emergentes como o desempenho e a expansão. Na modelagem são considerados os aspectos de objetos e os eventos que representam o comportamento esperado do sistema. No GeoMedicina é empregado o modelo entidade-relacionamento para a construção de um esquema relacional.

### 3.3.1 Dados Geoquímicos

Os dados geoquímicos são obtidos pela empresa Minerais do Paraná (MINEROPAR) e consistem em uma amostragem planejada de modo a proporcionar uma completa cobertura do território paranaense, que compreende cerca de 200.000 km<sup>2</sup>, com área amostrada correspondendo a 83% do Estado do Paraná. As amostras são coletadas nos canais dos rios, no local de maior fluxo d'água onde os sedimentos e a água sofrem constante homogeneização.

Para a representação das superfícies geoquímicas complexas, que definem as tendências geoquímicas dos elementos aqui considerados, é necessário realizar a regularização das malhas de amostragem. Essa regularização diz respeito ao espaçamento e a distribuição espacial dos dados originais. O ponto de amostragem foi deslocado de sua posição real para o baricentro da microbacia. Esse procedimento é adotado considerando que o ponto de coleta da amostra não representa apenas um ponto mais sim uma área (a microbacia) e que os pontos originais estão concentrados ao longo das linhas das drenagens principais [50]. Após este processo, os baricentros das bacias, através de um processo de interpolação, passam a constituir uma malha regular, baseada na *Global Geochemical Reference Network* [56].

### 3.3.2 Dados do Sistema de Saúde

Os dados do Sistema de Informações de Saúde são obtidos através da Secretaria de Saúde do Estado do Paraná (SESA), Seção de Vigilância Epidemiológica [94] e do Departamento de Informática do Sistema Único de Saúde (SUS) [22]. Estes dados correspondem ao número de mortalidade e morbidade por doença, segundo as classificações internacionais de doenças CID-9 (anterior a 1996) e CID-10 (após 1996), ocorrida nos municípios do Estado. Os dados disponíveis são oriundos do Sistema de Informações sobre Mortalidade (SIM), sistema este gerido pelo Departamento de Análise de Situação de Saúde, da Secretaria de Vigilância em Saúde [103], em conjunto com as Secretarias Estaduais e Municipais de Saúde.

Como há um atraso na publicação nacional destas informações devido à dependência das atualizações de todas as secretarias estaduais, dados recentes são obtidos diretamente na SESA.

Esses dados são fornecidos em formato de planilhas ou arquivos separados por vírgula. Uma importação dos mesmos para a base de dados do sistema é realizada com o uso de ferramentas do SGBD. Também, alguma manipulação via linguagem SQL é feita para que os dados sejam inseridos na base de dados espaciais.

No GeoMedicina são monitoradas doenças como o câncer infantil do córtex da supra-renal, as hepatites dos tipos B e C e as neoplasias constantes no capítulo II da CID-10.

### 3.3.3 Dados Cartográficos

Os dados que constituem a base cartográfica com os limites municipais e microrregionais são obtidos através do Sistema de Mapas Interativos do Instituto Brasileiro de Geografia e Estatística (IBGE) [43], onde se seleciona o tema Mapa de Divisões Territoriais. Estes dados, que originalmente estão no formato *ShapeFile*, são convertidos, por meio de um conversor do PostGIS, para uma base de dados espaciais.

No projeto GeoMedicina, o objeto geográfico complexo Estado consiste em objetos geográficos complexos denominados Regionais de Saúde, os quais são compostos por objetos geográficos atômicos, os Municípios. Embora esta hierarquia sugira o emprego do modelo topológico, neste projeto optou-se pelo modelo *spaghetti* devido a uma limitação do SGBD empregado que atualmente não suporta o modelo topológico. Assim, as computações de topologia devem ser feitas através de linguagens como o SQL. Para compensar esta limitação, pode-se pré-computar os relacionamentos topológicos necessários e armazená-los na base de dados.

Os dados geográficos são produzidos por organizações públicas e privadas, sendo defendida pela comunidade de usuários de FOSS4G a idéia de que os dados públicos sejam fornecidos sem custo, via Internet. No Brasil, algumas iniciativas pioneiras, lideradas pelo Ministério do Meio Ambiente, merecem destaque pela inovação no fornecimento de dados espaciais através do WMS [59]. Instituições como Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) - Monitoramento por Satélite [30] e Agência Nacional de Águas [1] fornecem seus dados espaciais segundo o padrão WMS. No exterior, estes padrões são fornecidos por organizações como NASA [62] e NOAA [65]. Uma lista atual dos serviços internacionais WMS pode ser

encontrada na página da Internet da empresa *Skylab Mobilesystems Ltd* [98].

### 3.4 Requisitos da Interface com o Usuário

No GeoMedicina, a interação dos usuários é implementada com o emprego dos três primeiros modos propostos por Shneiderman [97]. O estilo de preenchimento de formulário associado ao estilo de seleção de menu é empregado para que o usuário selecione as variáveis que deseja e controle a funcionalidade do sistema. O estilo de manipulação direta permite ao usuário interagir diretamente com o mapa na tela, deslocando-o por meio de cliques e arrastos do *mouse*.

A interatividade do usuário no GeoMedicina com a API Ka-map permite que ele modifique a visualização através da seleção de camadas sobrepostas no mapa digital, do deslocamento e da ampliação deste mapa sobre certas áreas. Também é possível salvar e imprimir as imagens digitais obtidas, possibilitando a inclusão das mesmas em documentos ou apresentações.

Dois perfis de usuários potenciais do sistema são considerados. O primeiro, pesquisador de epidemiologia que faz ensaios com as múltiplas combinações no sentido de descobrir padrões que indiquem os relacionamentos importantes entre doenças e fatores ambientais. Neste processo são identificadas as supostas áreas de risco à saúde humana que, no futuro, poderão ser validadas pela comunidade científica. O segundo usuário é o agente de saúde pública que tem o auxílio do sistema no processo de tomada de decisão para a alocação de recursos regionais destinados aos devidos cuidados com a saúde da população.

Ao visualizar e navegar em mapas temáticos, o usuário obtém a indicação dos lugares onde, para o tema em questão, há deficiências ou há condições ideais. Esta indicação, visualizada na sobreposição de camadas, auxilia-o na formulação de hipóteses sobre relacionamentos importantes entre fatores ambientais e determinadas doenças.

### 3.5 Protótipo Inicial - API Ka-map

Tendo em mente a funcionalidade esperada do sistema e a política de reuso de software, escolheu-se a API Ka-map, cuja arquitetura em camadas do protótipo inicial é apresentada na

Figura 3.1.

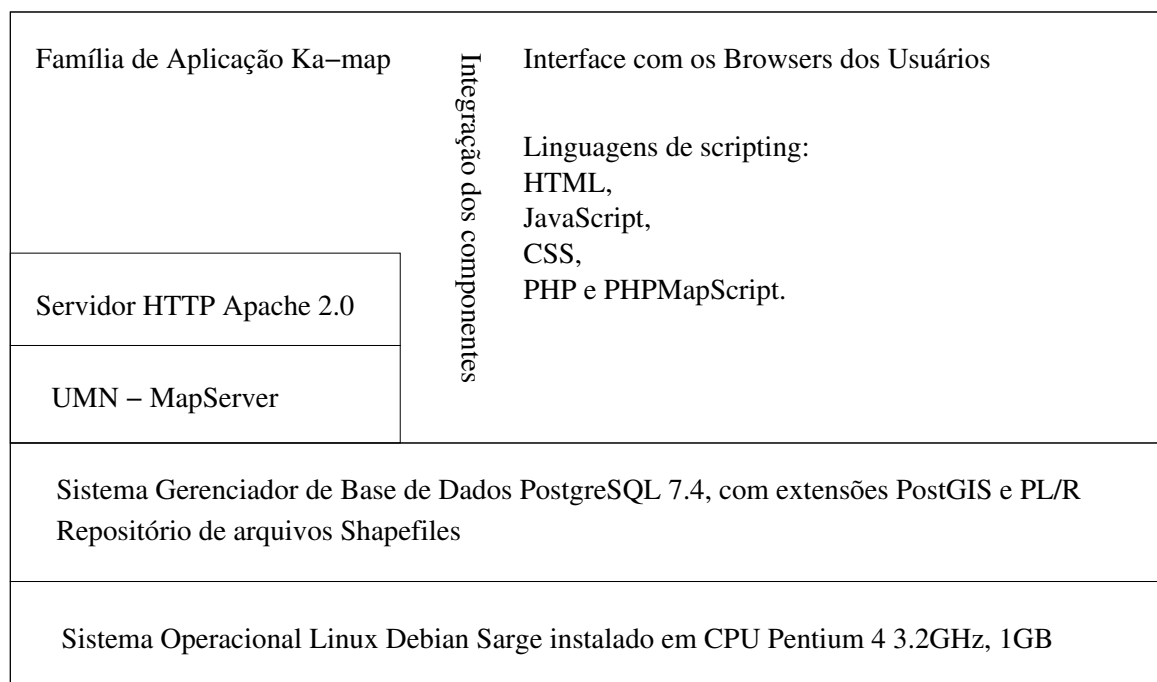


Figura 3.1: Arquitetura inicial do protótipo GeoMedicina.

Nas próximas seções são descritas as características mais importantes para este projeto dos softwares que compõem o sistema em desenvolvimento.

### 3.5.1 Servidor HTTP Apache

O sistema baseia-se na arquitetura cliente-servidor, onde o cliente é o navegador (*browser*) do usuário e o servidor HTTP é o Apache [3]. O cliente envia para o servidor requisições de conteúdos constantes em um formulário HTML, onde estão embutidos parâmetros que definem o que o usuário deseja obter. O servidor trata esta requisição e retorna ao cliente o conteúdo solicitado.

O conteúdo respondido pelo servidor HTTP consiste no envio de arquivos no formato HTML, conforme os itens selecionados em formulários submetidos pelo cliente. Este conteúdo torna-se dinâmico quando o HTML, ao invés de ser um arquivo estático, é gerado com a execução de *scripts* e programas CGI, os quais geram respectivamente, o código HTML e as imagens, conforme as seleções feitas pelo cliente.

Morimoto [61] explica que a principal característica do Apache é a modularidade. Ao invés de ser um aplicativo complexo, que desempenha muitas funções, o Apache se limita a executar uma única tarefa: entregar páginas HTML e outros tipos de arquivos aos clientes. Outras funções, como a interpretação de linguagens e o acesso ao SGBD, são executadas por módulos externos carregados sob demanda.

### 3.5.2 SGBD PostgreSQL e PostGIS

O PostgreSQL [84] é um SGBD objeto-relacional de código aberto, o qual, neste projeto, é estendido pelo PostGIS [83], passando a apresentar a habilidade de um gerenciador de base de dados espaciais [58]. Isto é possível devido à escalabilidade do PostgreSQL, o que permite estendê-lo, adicionando-lhe novos tipos de dados, funções, operadores, métodos de índice e linguagens procedurais. Assim, é possível embutir programas no SGBD escritos em várias linguagens além do SQL e C, como o PL/PgSQL, Perl, Python, Ruby e PL/R.

Outra característica importante do PostgreSQL é o recurso do emprego de expressões regulares nas cláusulas *where*. Isto, ao mesmo tempo em que simplifica as expressões usadas na procura por padrões, confere-lhes uma sofisticação que aumenta o poder de expressão e a velocidade de busca [26].

O PostgreSQL possui tipos geométricos, operadores espaciais e índices somente para retângulos e círculos. Pontos, linhas e polígonos podem ser indexados através de seus retângulos envolventes com o emprego de árvores R, mas isto implica em perda de precisão da forma destes objetos, degradando a capacidade de resolução espacial de funções como distância e intercessão. A solução para este problema é encontrada na extensão deste SGBD, o PostGIS. O índice usado nesta extensão é o GiST cuja aplicação, além de preservar a resolução espacial, não limita a indexação a *features* com tamanho inferior a 8 Kbytes, como é o caso do índice nativo do PostgreSQL.

O PostGIS, além de armazenar, tem a capacidade de manipular dados espaciais [58]. Esta característica contribui para a economia de banda de rede e de memória primária ou secundária. Caso não houvesse esta capacidade, os dados espaciais seriam transmitidos via rede e armazenados em memória para o posterior processamento via aplicação. Nesta extensão

está implementado o padrão *Simple Feature Specification for SQL (SFSQL)* [68] da OGC, o que assegura portabilidade ao longo do tempo de vida do projeto com o emprego de funções padronizadas e independentes da aplicação.

### 3.5.3 UMN-MapServer

O UMN-MapServer [108], mantido pela Universidade de Minnesota (UMN), é o componente integrador da família de aplicações que constitui o protótipo inicial do GeoMedicina. Ele integra softwares da camada de processamento de dados, que realizam o tratamento dos dados espaciais e que produzem as imagens digitais, com o servidor HTTP da camada serviço de dados. No contexto de funcionamento do MapServer, há dois componentes principais configurados no arquivo *MapFile* [54]: a fonte de dados espaciais (por exemplo, ShapeFile, base de dados) e as imagens de mapas em formatos legíveis por *browsers* (por exemplo, JPG, GIF e PNG).

A funcionalidade do MapServer consiste em coordenar a leitura e a manipulação de dados espaciais e em formatar os resultados do processamento entregando-os às ferramentas de produção dos mapas digitais. O controle e a configuração das entradas e saídas são pelo *MapFile*. Neste são declarados os formatos dos dados de entrada e das imagens produzidas, como, por exemplo: o estilo de apresentação das camadas e suas respectivas classes, o tipo de imagem de saída, os sistemas de projeção dos dados espaciais e do mapa digital gerado, o mapa de referência, a legenda, a barra de escala e o formato da apresentação das consultas espaciais. O MapServer é uma ferramenta de apresentação de informações sobre mapas por meio do uso de várias técnicas cartográficas para visualização de resultados.

As características mais importantes do MapServer para este projeto são: a automação dos elementos de mapas, como barra de escala, mapa de referência e legenda; a produção de mapas temáticos com o uso de expressões lógicas ou expressões regulares baseadas em classes; a projeção dinâmica de mapas. O MapServer é usado para desenhar os mapas vetoriais armazenados no *PostGIS* ou em *ShapeFiles*. As consultas SQL são expressas nos respectivos *layers* declarados no *MapFile*.

O MapServer é integrado ao servidor HTTP através das interfaces *Common Gateway*

*Interface (CGI)* [14] ou linguagens de *scripting* como PHP, Python, Perl ou Java. O CGI consiste em uma interface entre programas executáveis e servidores HTTP. Através desta interface, o programa em CGI recebe parâmetros de configuração e de modo de funcionamento, executa um processamento correspondente e retorna o resultado do seu processamento ao servidor HTTP.

### 3.5.4 PHP e Scripting

O PHP [78] é uma linguagem de *scripting* própria para o desenvolvimento de interfaces *Web*. Com o PHP é possível fazer com que uma página de conteúdo HTTP apresente um comportamento dinâmico.

No projeto GeoMedicina, o PHP é empregado como linguagem de *scripting* por meio de uma extensão chamada PHPMapscript [80], a qual é usada para a criação de uma API MapScript [53]. Esta API implementa uma interação dinâmica com o MapServer através do *MapFile*, o qual é transformado em classes [79] que podem ser acessadas como objetos, com atributos e métodos que atuam sobre os elementos do *MapFile*.

### 3.5.5 AJAX, JavaScript e XML

O *Asynchronous JavaScript and XML (AJAX)* [74, 101] é um conceito de desenvolvimento de conteúdo para Internet no qual se economiza o tráfego na rede, pois, com este conceito, não é necessário que a atualização de determinadas informações implique na modificação da página inteira. O AJAX proporciona um mecanismo para separar o dado da aplicação, fazendo com que somente o dado trafegue pela rede e somente quando for necessário para a aplicação. O AJAX diminui as chamadas ao servidor de mapas e o processamento realizado no servidor. Esta economia é alcançada com um sistema de subdivisão da imagem digital do mapa em retângulos pequenos (*tiles*), armazenados em *cache* do cliente e do servidor, o que permite que sejam requeridos pelo cliente somente os *tiles* ausentes no *cache* local e; também, que sejam processados pelo servidor somente as partes da imagem digital que ainda não constem no *cache* do servidor.

O balanceamento do processamento no lado do cliente com o processamento no lado do



servidor faz com que a interface com o usuário seja fluida e contínua. A fluidez e continuidade da interface são obtidas com o AJAX, como o próprio nome indica, a comunicação é assíncrona porque o *browser* não pára de mostrar o conteúdo da página enquanto espera que a requisição feita seja respondida. Os *scripts* do cliente usam o objeto JavaScript XMLHttpRequest para fazer chamadas ao servidor, o qual envia as respostas em XML que são analisadas em segundo plano pelo cliente.

O JavaScript monitora e controla os eventos usados para a interação do usuário com o sistema. Assim, um evento como clicar e arrastar o *mouse* para deslocar o mapa é tratado pela aplicação. Nesta ação, através de controles, pode-se evitar uma requisição ao servidor, recuperando-se as partes do mapa que foram visualizadas em navegações anteriores constantes na memória *cache* do cliente.

Uma parte do JavaScript que está rodando no *browser* cria uma instância do objeto XMLHttpRequest e uma função que serve como um *callback* assíncrono. O *script* então usa o objeto XMLHttpRequest para dirigir uma operação HTTP ao servidor. Quando a resposta é recebida, a função *callback* é invocada. O dado retornado pelo servidor é processado dentro da função *callback*. Se o dado está codificado em XML, o objeto XMLHttpRequest analisará o dado automaticamente usando os mecanismos de processamento XML integrados ao *browser* [100]. Além do JavaScript e do XML o AJAX emprega as folhas de estilo em cascata e o modelo de objeto de documento, vistos a seguir.

Algumas características de segurança que envolvem o procedimento de invocar serviços Web através do JavaScript embutido em um *browser* devem ser entendidas. Há uma restrição nos *browsers* quanto à interação, autorizada somente para serviços hospedados no mesmo domínio da página Web. Para habilitar o acesso a serviços Web localizados em outro domínio, uma solução é passar todas as requisições do XMLHttpRequest através de um *proxy* localizado no mesmo domínio que a página carregada. Este proxy encaminha as requisições para a localização remota e retorna os resultados ao *browser*. Assim, sob o ponto de vista do XMLHttpRequest, a interação ocorre no mesmo domínio da página.

### 3.5.5.1 Cascading Style Sheets - CSS

As folhas de estilo em cascata são um mecanismo simples para a adição de estilo nas páginas *Web*. As folhas de estilo descrevem como documentos são apresentados na tela ou em impressoras através da definição de, por exemplo, fontes, cores e espaçamento [111]. O nível dois desta especificação [112] permite que os *designers* incluam estilo a documentos estruturados, sejam em documentos HTML ou em aplicações XML.

Com o emprego das folhas de estilo separa-se o estilo de apresentação do conteúdo desses documentos. Isto uniformiza o *design* de uma aplicação *Web* e facilita a manutenção do seu conteúdo. O uso das CSS consiste em incluir estilos de títulos, barras de ferramentas, menus e tabelas definidos uma única vez nos documentos. A expressão cascata refere-se à capacidade de mais de um estilo influenciar a apresentação simultaneamente.

### 3.5.5.2 Document Object Model - DOM

O modelo de objeto de documento, do inglês *Document Object Model* (DOM), é uma interface de programação de aplicativos independente de plataforma ou linguagem que possibilita aos programas e *scripts* acessarem e atualizarem dinamicamente o conteúdo, o estilo e a estrutura de documentos [113].

O objetivo do DOM é facilitar aos programadores o acesso aos componentes de um documento, viabilizando a eliminação, a adição e a modificação do conteúdo, atributos e estilo destes componentes estruturados. No DOM, as *tags* de um documento HTML são estruturadas em uma árvore onde o nó raiz corresponde a *tag HTML*. Abaixo da *tag BODY* estão, por exemplo, componentes como formulários, tabelas, caixas de texto, botões, listas, etc. O AJAX usa esta árvore onde estão incluídos também os estilos de apresentação para manipular o conteúdo de uma página *Web* conforme a interação dos usuários controlada pela interface do sistema.

### 3.5.6 Escolha dos Componentes de Software

Algumas vantagens que influenciaram a escolha dos componentes de software já foram expostas ao longo deste texto, como o comportamento contínuo de uma interface AJAX sob a qual foi implementada a API Ka-map.

O servidor de mapas MapServer e a extensão PostGIS do SGBD possuem a capacidade de interoperação com serviços de mapas na *Web*. A experimentação de outros sistemas de mapas na *Web* mostrou que as soluções que empregaram tecnologia similar à escolhida são tão eficientes quanto o protótipo do GeoMedicina.

A argumentação dos próximos parágrafos tem o objetivo de mostrar que os componentes escolhidos apresentam um desempenho equivalente a softwares comerciais empregados atualmente em *Web mapping*.

Uma opção ao MapServer seria o ESRI-ArcIMS [4]. Anderson [2] da *Refractions Research* comparou estes servidores de mapas em testes descrevendo aspectos como a facilidade de administração, a conformidade com o padrão WMS 1.1.1, a interoperação com outros softwares e o desempenho. A base de dados usada nos testes foi gerenciada pelo ESRI-ArcSDE 8.3 [5] e as requisições ao servidor consistiram em *GetMaps* criadas em uma extensão do JMeter [45].

Nos primeiros testes de desempenho, onde se comparou o número de requisições concorrentes contra o tempo de resposta, constatou-se a superioridade do ArcIMS. Verificou-se que o desempenho baixo do MapServer deveu-se a dois principais gargalos: o custo de 1 a 2 segundos para cada requisição *GetMap* mais o custo de extração das *features* do ArcSDE. A solução para diminuir estes custos consistiu em prover conexões persistentes com a base de dados. Para isso foi necessário adicionar o suporte *FastCGI* [33], uma vez que não há meios de manter conexões persistentes com o MapServer no modo CGI. Também foi necessário melhorar o código do MapServer que obtém as *features* do ArcSDE, diminuindo o uso das operações mais custosas. A partir destas melhorias, os primeiros testes foram repetidos constatando-se um desempenho equivalente para ambos os softwares.

Outros testes de desempenho foram feitos com a medição do tempo de resposta na geração de imagens em formatos GIF, PNG e JPG e na reprojeção de mapas. Constatou-se que o desempenho para imagens do tipo JPG é melhor com o ArcIMS, enquanto que para imagens

GIF e PNG o desempenho é superior com o MapServer. Este também apresentou um desempenho melhor na reprojeção de mapas e foi mais rápido que o ArcIMS no teste da quantidade de mapas servidos por segundo com acréscimo no número de usuários concorrentes sobre um período de tempo estendido. O tempo de resposta na extração de *features* também foi inferior no MapServer.

Quanto à conformidade com o padrão WMS, o MapServer passou em todos os 83 testes realizados enquanto que o ArcIMS passou em 71, sendo as 12 falhas atribuídas a algumas respostas com tipo MIME errado e algumas exceções sinalizadas com código errado. Quanto à facilidade de administração, o MapServer apresentou as vantagens de reiniciar tão rápido quanto o servidor *Web* e de não ser necessário recarregar os serviços quando há alguma mudança nos arquivos de configuração. O ArcIMS apresentou a vantagem de um controle mais granular sobre os níveis de *log*.

Segundo as investigações da Netcraft [106], o Apache é o servidor *Web* mais usado entre todos os *sites* ativos em março de 2007, com 58,62% de servidores Apache contra 31,02% dos servidores Microsoft. Questões sobre o desempenho destes produtos apontam vantagens tanto para o IIS 6 da Microsoft quanto para o Apache 2. Um argumento convincente na escolha do produto refere-se ao ambiente no qual este é integrado. Se o servidor HTTP for integrado a softwares da Microsoft é melhor usar o IIS, mas se o servidor HTTP for integrado a softwares livres é melhor usar o Apache [19]. O baixo custo na aquisição do software livre, o que é o caso de todos os componentes da aplicação, também é uma vantagem levada em conta.

### 3.5.7 Instalação e Configuração do Protótipo Inicial

Os componentes de software livre foram instalados em um microcomputador padrão com a seguinte configuração: processador de 3.2 GHz, memória primária de 1 GBytes, memória secundária *Small Computer Systems Interface (SCSI)* de 182 GBytes e sistema operacional Linux Debian [25] Sarge 3.3.5, kernel 2.4.27.

O custo dessa máquina é relativamente baixo quando comparado ao custo de servidores profissionais. O uso de um microcomputador padrão é possível graças a uma versão do sistema

operacional Linux desenvolvida especificamente para esta arquitetura de computador. Além do baixo custo de hardware é importante ressaltar que o custo com a aquisição de software é praticamente nulo, o que consiste em uma economia significativa. O procedimento para a instalação dos pacotes é descrito no Apêndice A.

A camada superior do protótipo inicial é a API Ka-map, a qual é constituída por *scripts* PHP e JavaScript, folhas de estilo em cascata (CSS) e um mínimo de hipertexto estático. Estes arquivos constituem a aplicação localizada na árvore do serviço *Web*, sendo o `"/var/www/"` a opção *default* para o Apache obtido em repositório Debian. Na mesma árvore de diretórios do código fonte da API, há um conjunto de dados espaciais em ShapeFiles e uma configuração em MapFile para fins de demonstração. O primeiro teste da aplicação é feito com este conjunto de dados, sendo necessária a intervenção do programador somente para o ajuste do nome de biblioteca dinâmica PHPMapscript, caso necessário. Outra localização que deve ser definida é a do *cache* do Ka-map.

Depois de testado o funcionamento do protótipo com os dados da demonstração, é necessário adquirir os dados previstos para o projeto e editar o *MapFile* para a inclusão da configuração destes dados. No GeoMedicina, estes dados iniciais são *ShapeFiles* de polígonos que, quando dispostos e coloridos conforme as classes de teores de elementos químicos, formam mapas de isolinhas. Outros mapas com a razão de doenças por 100 mil habitantes de cada município são adicionados. Um procedimento de importação destes dados em formato DBF para a base de dados foi criado com o uso da linguagem SQL. O Apêndice B apresenta os comandos usados nesta importação.

O *MapFile* do Ka-map foi modificado para a leitura dessas fontes de dados espaciais e a criação das imagens mostradas na tela. O Apêndice C apresenta trechos do *MapFile* modificado. Algumas telas do protótipo podem ser vistas no Capítulo 4.

### 3.5.8 Avaliação do Protótipo

Para o sucesso de um projeto de sistema dedicado à pesquisa, em que profissionais com diferentes formações são os usuários finais, é necessário um envolvimento de pessoal da organização, onde os mesmos participam do processo de concepção e aperfeiçoamento do sistema. Com

esta participação, a equipe de desenvolvimento projeta um sistema que atende às reais necessidades dos usuários. Além dos requisitos dos usuários, é necessário considerar os requisitos do sistema, através da pesquisa da tecnologia e metodologia modernas. A conformidade do sistema com o “estado da arte” da engenharia de software propicia a este um maior tempo de vida e a capacidade de expansão e aumenta a confiança do usuário no sistema.

A partir da experimentação do protótipo, os usuários o avaliam e o validam. Nestas avaliações surgem novos requisitos os quais propõem aprimoramentos ou mudanças nas propriedades funcionais e que, ao serem implementadas, constituirão uma nova versão. Esta nova versão é avaliada e validada no ciclo do modelo espiral. Uma forma de verificar a impressão que o usuário tem com a interface do sistema é a aplicação de um questionário, através do qual a equipe de desenvolvimento identifica o que precisa ser melhorado no sistema.

Neste trabalho, adaptou-se o *Questionnaire for User Interface Satisfaction* (QUIS) proposto por Shneiderman em 1987 e aperfeiçoado por pesquisadores da *University of Maryland*, o qual tem a finalidade de verificar:

- o tempo que o usuário demora em assimilar uma funcionalidade do sistema;
- a velocidade com que as tarefas são executadas;
- a suscetibilidade a erros;
- o tempo de retenção de comandos pelo usuário;
- a satisfação subjetiva do usuário.

### **3.6 Projeto de Evolução e Manutenção da Aplicação**

Considerando-se os objetivos do projeto, os trabalhos afins revisados e partindo-se do protótipo inicial, pretende-se um aprimoramento do sistema, onde novas funcionalidades são acrescentadas e interfaces são adaptadas. Para organizar o processo de software, é apresentado um conjunto de tarefas executados para atingir a evolução pretendida, descritas a seguir:

- Modelagem da base de dados espaciais usando os padrões descritos na seção 3.3.

- Implementação da metodologia e tecnologia estatística proposta na seção 2.6.
- Desenvolvimento de interface com o usuário de forma a atender aos conceitos apresentados nas seções 2.4 e 2.5.

Estas tarefas são decompostas em um conjunto de atividades descrito a seguir:

- modelagem da base de dados em um diagrama entidade-relacionamento;
- criação da base de dados via *scripts* SQL para a criação das tabelas, restrições e relacionamentos, a leitura de arquivos CSV e a inserção de dados;
- implementação de métodos de classificação estatísticos para a classificação automática dos temas apresentados ;
- implementação de métodos de suavização estatística para a expressão apropriada da distribuição espacial dos temas;
- implementação de consultas espaciais aos temas apresentados no mapa para a recuperação de atributos quantitativos ou qualitativos dos objetos geográficos;
- aperfeiçoamento do formulário *Web* com opções de seleção de doença e de elementos químicos para a geração de mapas temáticos;
- implementação de função gráfica que desenha retângulo sobre o mapa para a delimitação manual de áreas onde há hipótese de relacionamento entre elementos químicos e doença, usando linguagem de manipulação direta na interface;
- implementação de inserção de polígono definido pelo usuário na base de dados, na tabela que contém o mapa do usuário;
- implementação de controle de acesso a usuários através de *login* com senha e níveis de permissão com o estabelecimento de três níveis de acesso à informação: hipótese (acessado somente pelo pesquisador), verificação (acessada pela rede de pesquisadores), e demonstrado (acessado pelos pesquisadores e agentes da saúde pública).

### 3.6.1 Rede de tarefas

Uma rede de tarefas é usada para a distribuição e controle das atividades entre os membros da equipe de desenvolvimento. A Figura 3.2 apresenta esta distribuição, onde a interdependência entre as atividades está implícita na seqüência definida pelo início e pelo fim e o direcionamento das setas ao longo do diagrama.

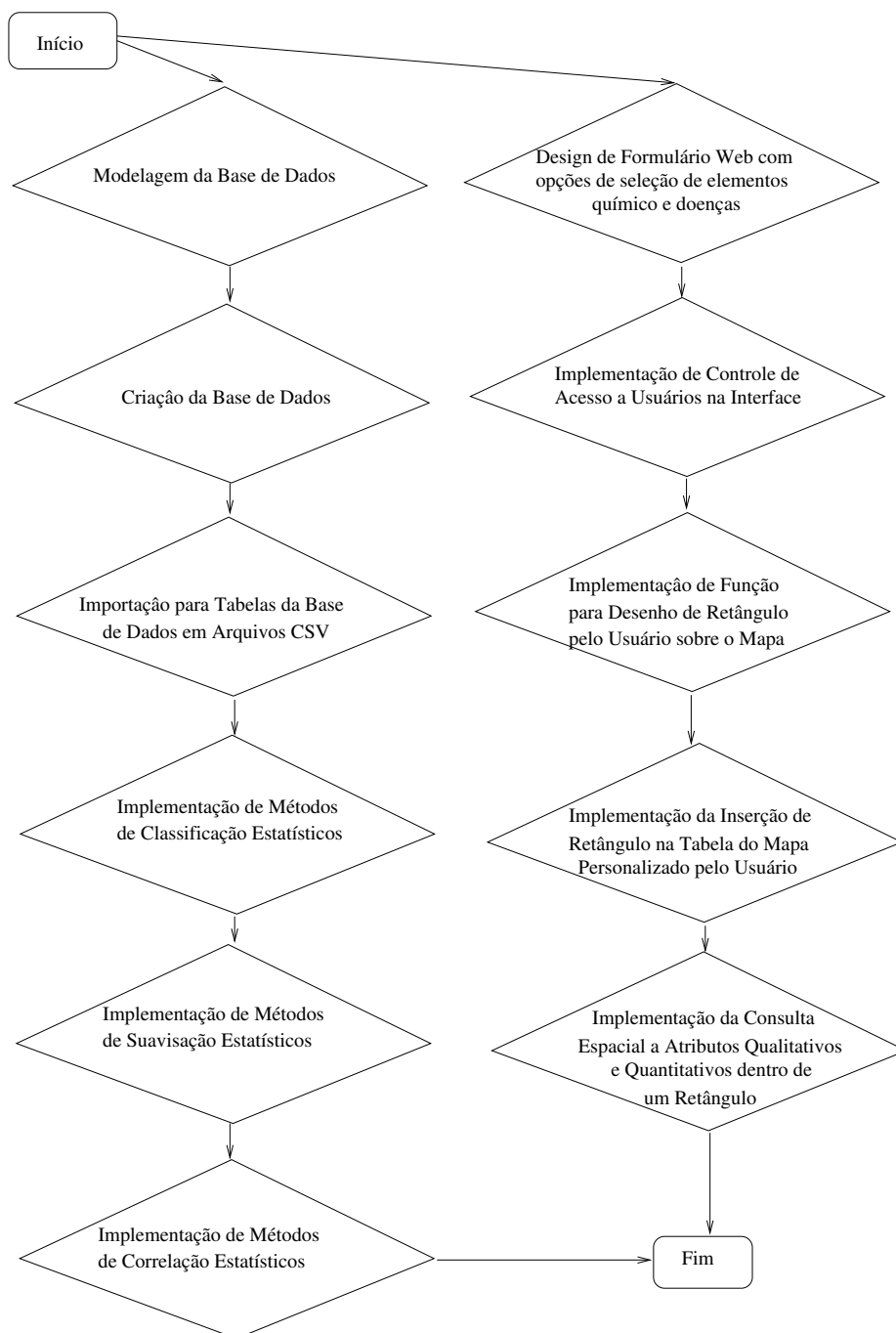


Figura 3.2: Rede de tarefas do projeto de evolução do protótipo.



### 3.6.2 Diagrama Entidade-Relacionamento

O diagrama entidade-relacionamento (ER) da Figura 3.3 apresenta as entidades da base de dados espacial do GeoMedicina. Os limites administrativos dos municípios são definidos em polígonos na entidade MUNICIPIO. O atributo “gid” é a chave primária com um índice de árvore B. Há um segundo índice espacial do tipo GiST, criado para o atributo geoPoligono. A identificação do município (IdentMun) e o período em que são válidos os limites municipais (anoIni e anoFim) são usados para relacionamentos com outras entidades cuja chave primária é a identificação do município associada ao período a que correspondem os demais atributos. A atomicidade do atributo temporal é o ano, o qual foi definido considerando-se a periodicidade em que as informações de saúde e geológicas são atualizadas. A grade geoquímica (GRADE-GEOQUIMICA) com os teores de elementos químicos em pontos interpolados; os indicadores de mortalidade por determinada doença nos municípios (MORTALIDADE) com a razão bruta da mortalidade por doença e os seus valores suavizados por métodos Bayesianos; os teores de elementos químicos nos municípios (TEORELEMQUIMMUN); os usuários cadastrados do sistema (USUARIOSISTEMA); e os mapas inseridos pelo usuário (MAPAUSUARIO) são as outras entidades constantes neste diagrama.

No diagrama ER, observa-se que os objetos geográficos atômicos, que são os polígonos com as fronteiras dos municípios e os pontos da grade regular geoquímica com os respectivos níveis de teor dos elementos químicos, relacionam-se entre si através da relação INTEGRA. Esta relação consiste em uma funcionalidade que percorre o conjunto de polígonos e obtém os pontos da grade que estão dentro de cada um dos polígonos. Estes pontos integrados ao respectivo município permitem que a relação APRESENTA calcule o teor médio de cada elemento químico o qual é armazenado no atributo teorMedio da entidade TEORELEMQUIMMUN.

### 3.6.3 Criação e Inserção na Base de Dados Espacial

O diagrama ER fornece uma visão do esquema da base de dados que auxilia na elaboração dos comandos para a criação da base e a inserção de dados nas tabelas. Um código na linguagem SQL foi escrito com os comandos que criam uma base preparada para a inserção

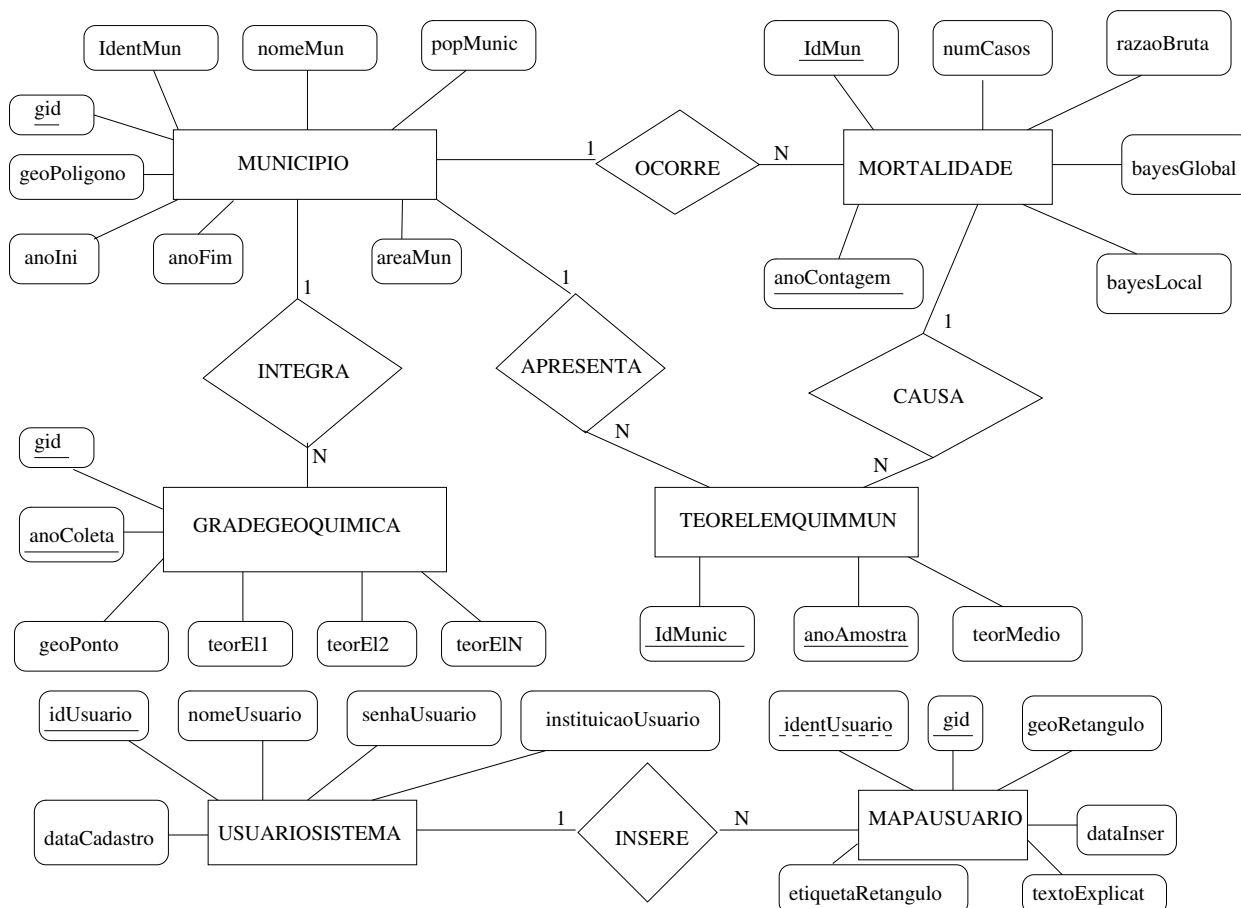


Figura 3.3: Diagrama entidade-relacionamento do GeoMedicina.

de objetos geográficos e seus atributos, incluindo as restrições e o relacionamento entre estes objetos. Um código na linguagem PHP foi escrito para ler arquivos em formato CSV e inserir registros na base criada. Um conversor de arquivos no formato *ShapeFile* para a base de dados PostGIS é usado para a inserção de registros nesta base. O Apêndice B apresenta os comandos executados e os códigos SQL escritos para esta finalidade, na seguinte seqüência:

- comando para a criação da base de dados com extensão espacial;
- conversão de *ShapeFile* com polígonos em tabela da base de dados criada;
- elaboração e execução de código em SQL com comandos de criação das tabelas, restrições e relacionamentos da base de dados; e
- elaboração de código SQL através do qual são feitas as leituras de arquivos CSV e as inserções de registros na base de dados criada.

## CAPÍTULO 4

### RESULTADOS OBTIDOS

Este capítulo apresenta os resultados dos experimentos efetuados com as versões do protótipo obtidas na execução do plano de evolução e manutenção da aplicação, onde a interface com o usuário é aperfeiçoada para atender às necessidades do sistema e aos requisitos planejados no projeto.

#### 4.1 Interface do Protótipo Inicial

A primeira versão do protótipo possui uma interface com o usuário apresentada no navegador *Web*, a qual é subdividida em área de apresentação cartográfica, em área da legenda, em mapa de referência e em barra de ferramentas. Os controles da aplicação posicionados nestas áreas são descritos a seguir:

- *Check boxes* na área da legenda que apresentam as camadas disponíveis para visualização e servem para selecionar quais temas (elemento químico, doença, base cartográfica) são mostrados na área de apresentação cartográfica.
- Botões associados às camadas na área da legenda que servem para regular a opacidade destas no mapa, permitindo a visualização de camadas inferiores por meio do aumento da transparência ou diminuição da opacidade de camadas superiores.
- Botões associados às camadas na área da legenda que servem para trocar a ordem das camadas, possibilitando que uma camada seja trocada de posição com a camada imediatamente inferior ou com a camada imediatamente superior.
- Botões associados às camadas na área da legenda que servem para mostrar ou ocultar a respectiva legenda.

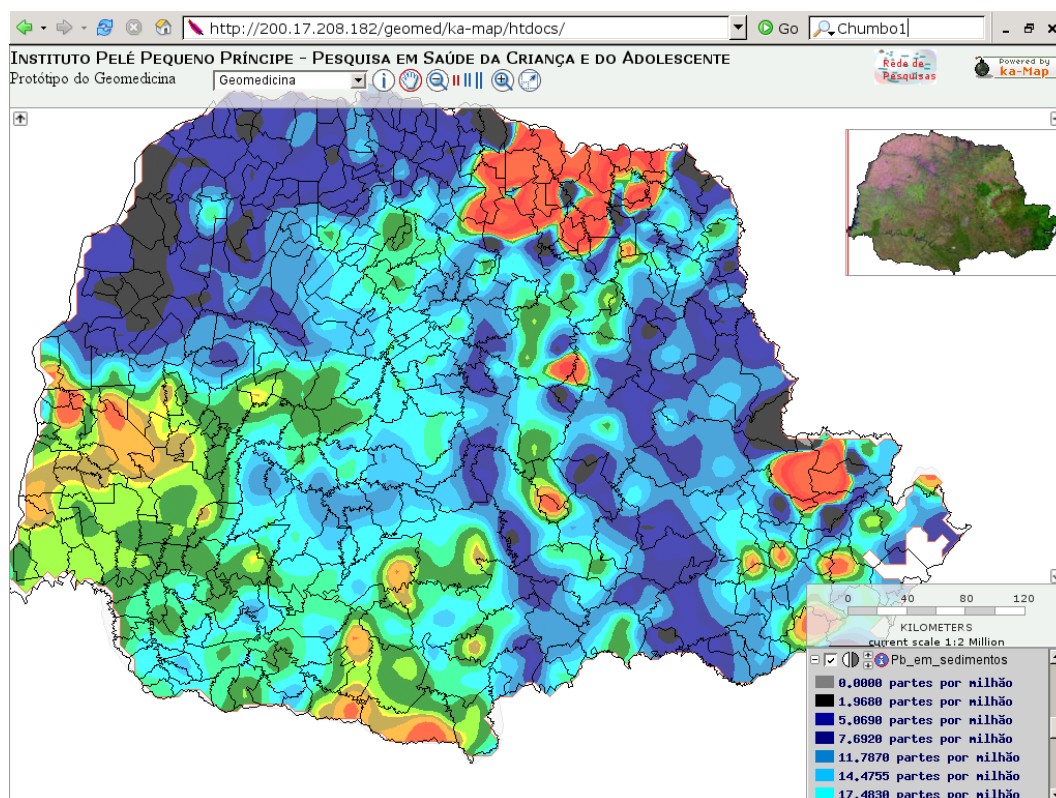
- Botões da barra de ferramentas que servem para modificar a escala do mapa, ampliando ou reduzindo a imagem digital visualizada.
- Botões na área de apresentação cartográfica que servem para apresentar ou ocultar as áreas sobrepostas à área de apresentação cartográfica. As áreas controladas com este recurso são a do mapa de referência e a da legenda.

Na área de visualização cartográfica, há interações com o usuário através de linguagem de manipulação direta, como por exemplo, o clicar e o arrastar do *mouse* para deslocar o mapa, ou o duplo clique sobre um ponto do mapa para centralizá-lo.

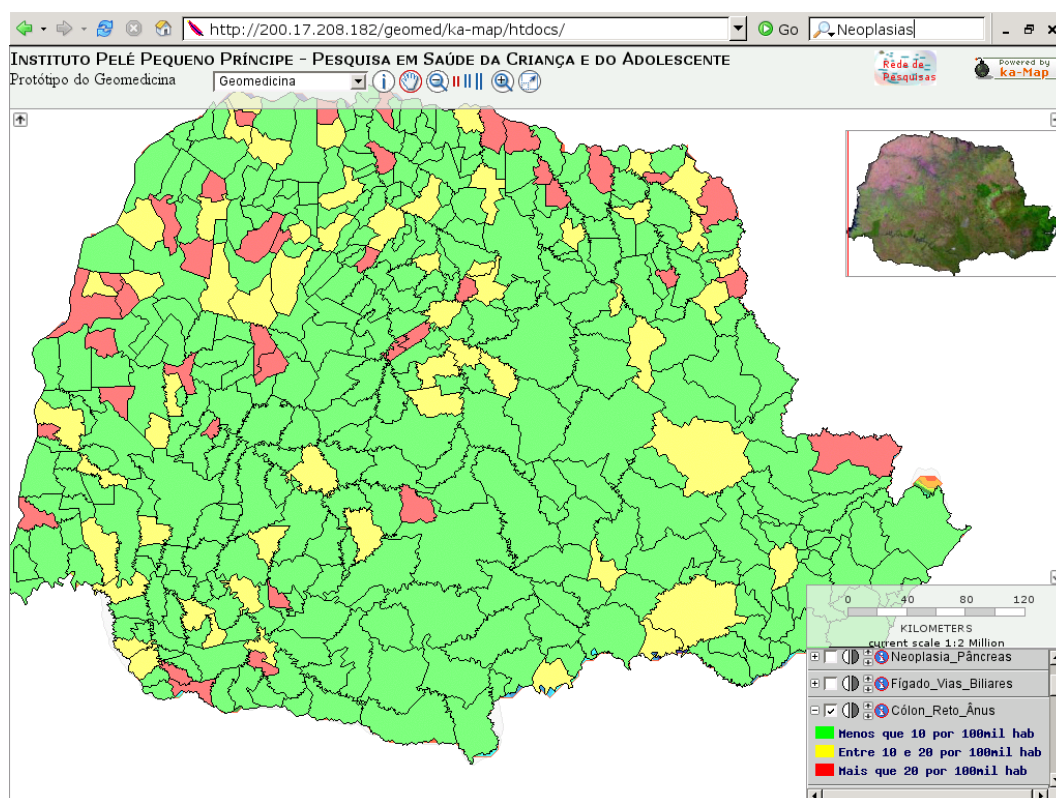
Um exemplo deste protótipo pode ser visto na figura 4.1. Esta figura mostra a distribuição espacial de elemento químico e de doença. A partir da legenda, pode-se verificar o intervalo de classe a que cada cor está associada. O teor do elemento químico é indicado na unidade partes por milhão e o coeficiente da doença, número de casos por cem mil habitantes.

A figura 4.2 apresenta uma sobreposição dos dois fenômenos, obtida através de um controle que modifica a transparência (opacidade) da camada superior.

Esta versão foi apresentada junto com o Plano de Trabalho, GEOMEDICINA: Áreas de Risco à Saúde Relacionadas com Fatores Ambientais, Constituição de uma Rede de Pesquisas e Implementação de um SIG, à Secretaria de Estado da Ciência, Tecnologia e Ensino Superior do Paraná (SETI) e aceito pela Unidade Gestora de Fundos no ano de 2006 [38].



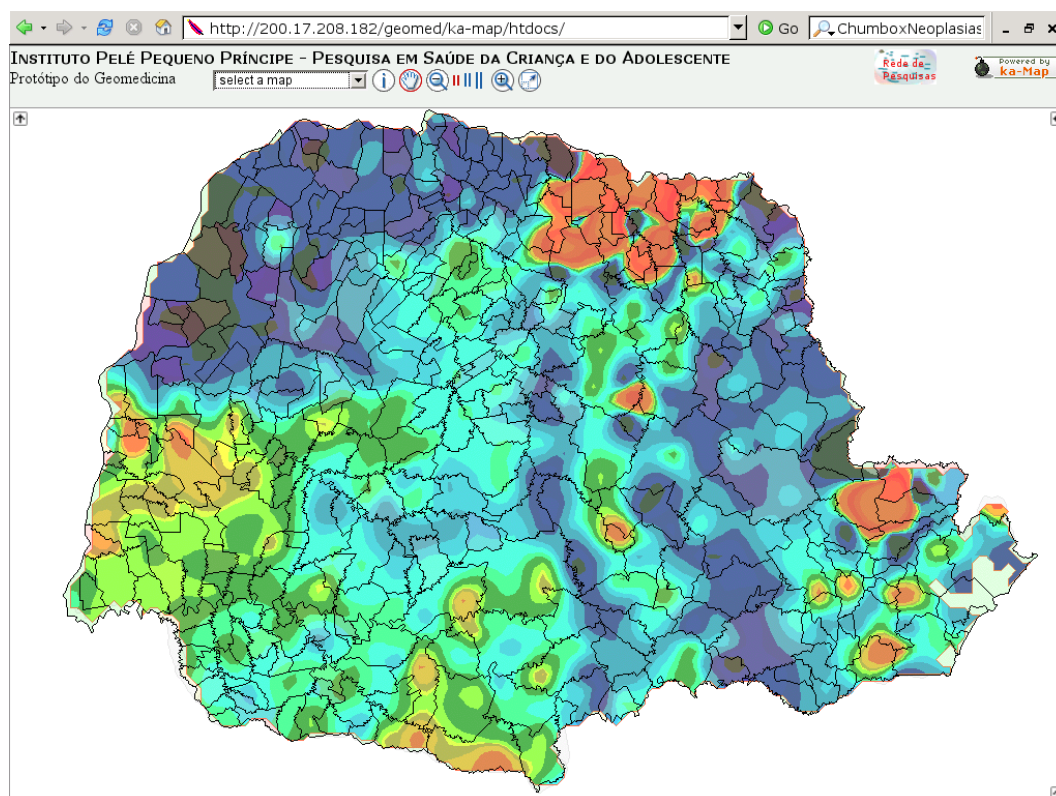
(a)



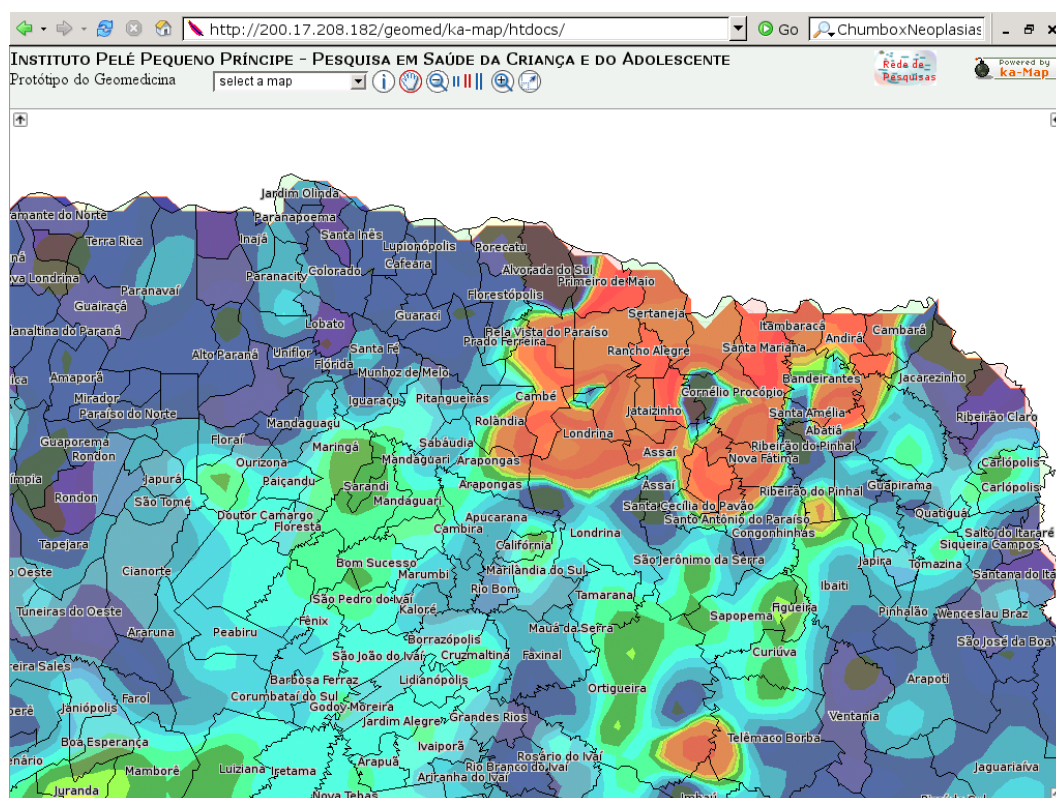
(b)

Figura 4.1: Protótipo inicial do GeoMedicina. Visualização geográfica de (a) Chumbo em sedimentos e de (b) Câncer de Colo de Intestino. Fontes: MINEROPAR 1996 e DATASUS 2004.





(a)



(b)

Figura 4.2: Protótipo inicial do GeoMedicina. Visualização geográfica da relação entre (a) Chumbo e Câncer de Colo de Intestino; (b) ampliação de (a). Fontes: MINEROPAR 1996 e DATASUS 2004.

## 4.2 Interface do Segundo Protótipo

Uma evolução do protótipo foi desenvolvida para a apresentação dos resultados iniciais da pesquisa sobre o Câncer de Córtex da Glândula Supra-renal [18]. Esta pesquisa motivou o projeto GeoMedicina, quando os pesquisadores procuraram descobrir se há padrões de localização da população que sofreu uma mutação genética do gene P53 e também quais são as variáveis do ambiente que influenciam esta mutação. Nessa pesquisa, foi comprovado que todas as crianças acometidas pela neoplasia da supra-renal apresentavam a referida mutação.

A contagem das mutações é realizada a cada teste positivo de DNA, o qual é feito sobre uma gota de sangue colhida em membrana apropriada nas maternidades do Estado do Paraná, semelhante ao “teste do pezinho” realizado em recém-nascidos. O laboratório do Instituto Pelé Pequeno Príncipe (IPPP) recebe das maternidades, para cada-recém nascido cuja mãe consentiu que fosse realizado o teste de DNA, um formulário, junto com a amostra do sangue, no qual constam dados como o município de residência da mãe, a etnologia familiar e um campo para o relato de câncer na família.

Digitadores do IPPP inserem esses dados em um sistema de informação, a partir do qual é possível verificar o número de testes realizados nos municípios de residência e o número de testes positivos. Estes valores são exportados para um arquivo CSV, o qual é usado para inserção dos dados em uma base de dados espacial.

Nesta segunda versão do protótipo, houve um melhoramento da interface com o aumento da área da legenda para onde foi remanejado o mapa de referência e onde foi incluída: uma caixa de texto para pesquisa de atributos e uma lista de seleção para a escolha do tipo de imagem a ser gravada. Estes elementos são alternados na área da legenda pelo acionamento de botões do tipo rádio na barra de ferramentas. Com o aumento desta área, a visibilidade das camadas e de suas respectivas legendas foi melhorada. O *layout* desta interface é apresentado na figura 4.3.

Além do aperfeiçoamento do *layout*, a partir da segunda versão do protótipo foram acrescentados ao sistema os seguintes recursos:

- Gravação no computador do usuário da imagem que está sendo visualizada na área de

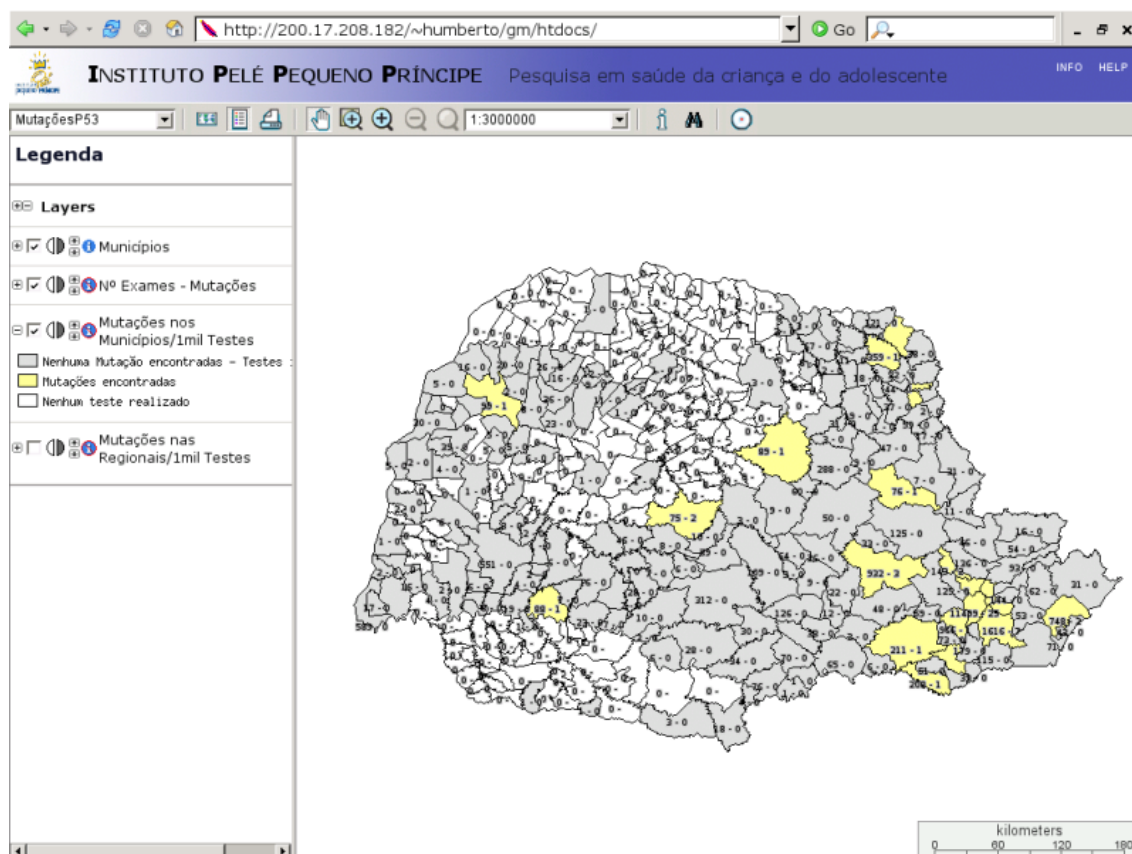


Figura 4.3: Layout da segunda versão do protótipo GeoMedicina. Mapa com os primeiros 30 mil testes de DNA realizados e mutações do gene P53 encontradas. Fonte: IPPP 2006.

apresentação cartográfica. Isto permite que os mapas sejam incluídos em documentos e apresentações. As figuras 4.4 e 4.5 são exemplos de imagens digitais obtidas com este recurso;

- Localização de objeto geográfico através da digitação de um de seus atributos em caixa de texto e do acionamento do botão “Procura”;
- Ampliação da imagem a partir de delimitação da área de interesse com um retângulo definido pelo usuário.

Foram apresentados apenas os primeiros dados coletados quando os testes de DNA no IPPP estavam no início (primeiros 30 mil testes) e ainda não havia uma amostragem significativa para a estatística. Assim, optou-se por produzir duas camadas no mapa da mutação do gene P53. Na primeira camada, apresentada na figura 4.3, constam os municípios do Estado do Paraná que apresentaram a mutação, incluindo etiquetas visíveis com o número de testes



realizados e o número de mutações verificadas. Nos municípios destacados em amarelo havia registro de uma ou mais mutações, nos municípios em cinza nenhum caso de mutação havia sido constatado e nos municípios em branco nenhum teste havia sido realizado. A figura 4.4 apresenta um exemplo de ampliação deste mapa.

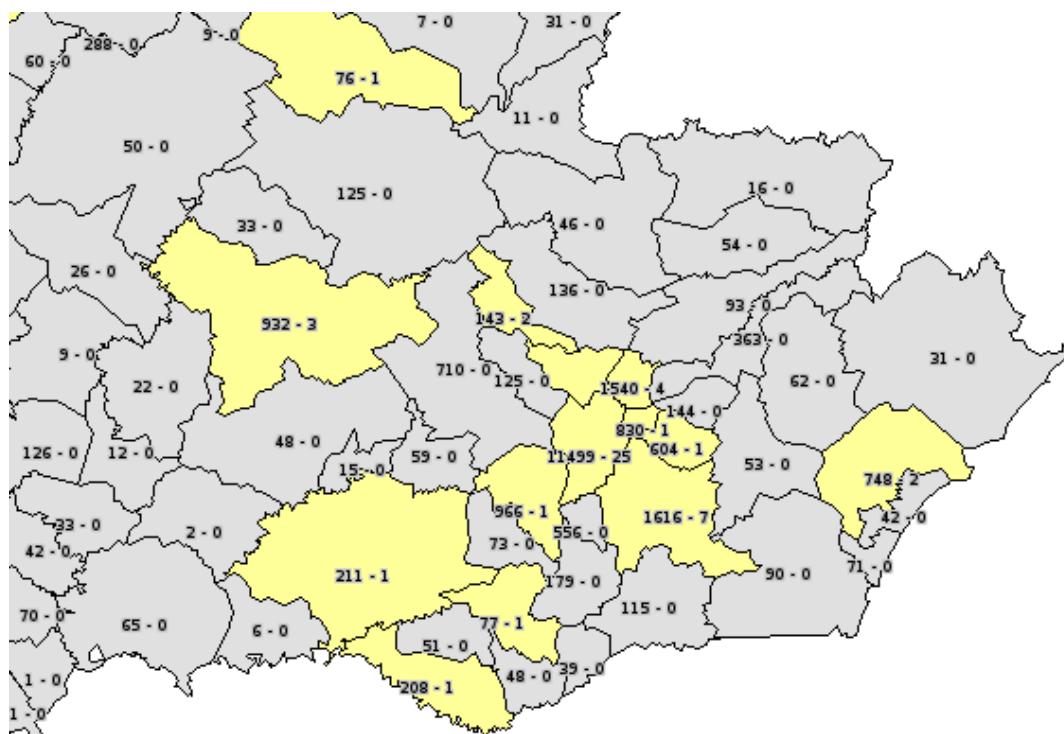


Figura 4.4: Ampliação do mapa com número de testes de DNA realizados e mutações encontradas. Fonte: IPPP 2006.

Em outra camada do mapa constam as razões entre o número de mutações encontradas por mil testes realizados nas regionais de saúde do Estado. Neste mapa, devido à maior abrangência, foi possível estabelecer uma classificação entre as razões obtidas. Os cálculos da razão foram feitos com a linguagem SQL. A figura 4.5 expressa em tons de marrom, do mais claro para o mais escuro, as regiões que apresentam, da menor à maior razão calculada. Etiquetas numéricas apresentam estas razões.

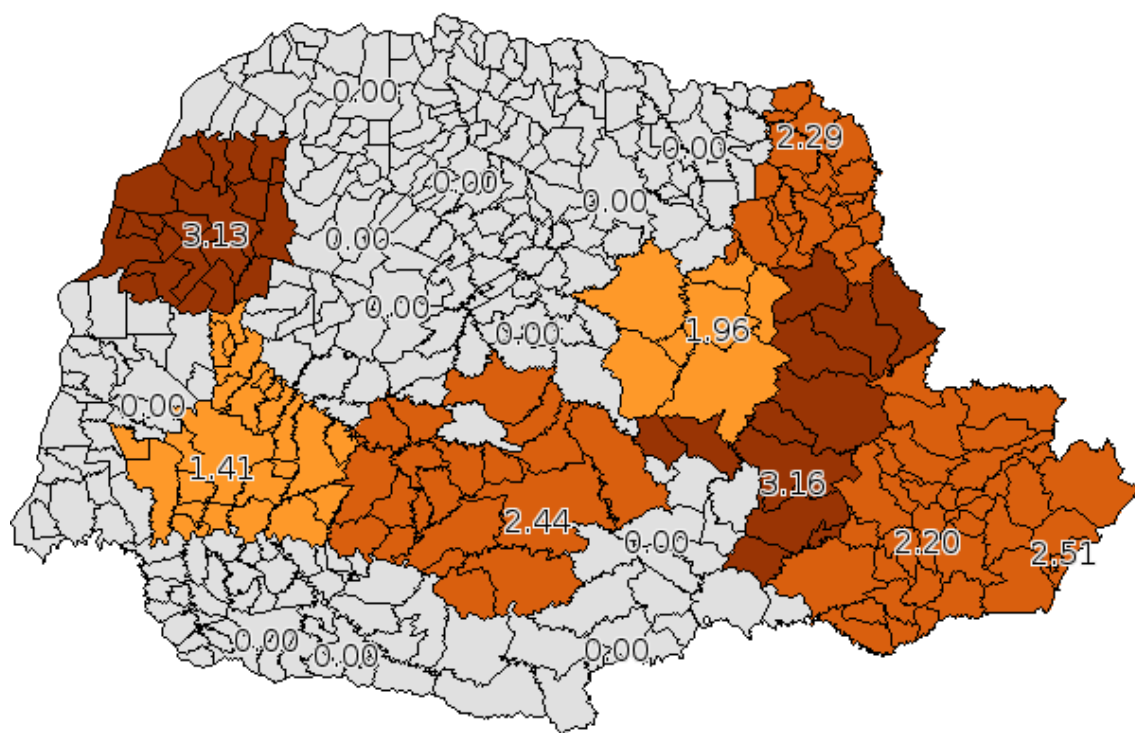


Figura 4.5: Mapa com uma classificação das mutações constatadas nas Regionais de Saúde. Fonte: IPPP 2006.

### 4.3 Interface do Terceiro Protótipo

Na terceira versão do protótipo foram integrados os temas constantes na primeira versão à interface constituída na segunda versão, além da integração de imagens satélite WMS. Estas imagens são obtidas do serviço de mapas do *Jet Propulsion Laboratory* (JPL) da NASA. A visualização destas imagens permite a identificação de cursos de água, bacias hidrográficas, aglomerados urbanos e uso do solo.

A integração de mais uma camada de software à arquitetura do GeoMedicina é apresentada na figura 4.6. As chamadas aos serviços de mapas, WMS, é implementada em código inserido no *MapFile*. Um exemplo de código para acesso a serviço WMS empregado no GeoMedicina é apresentado no Apêndice C.

Exemplos da integração dos temas podem ser vistos nas figuras 4.7 e 4.8, onde são apresentados, respectivamente, os teores do elemento químico Bário em águas e do elemento químico Chumbo em sedimentos de fundo dos rios. As camadas com os teores de elementos

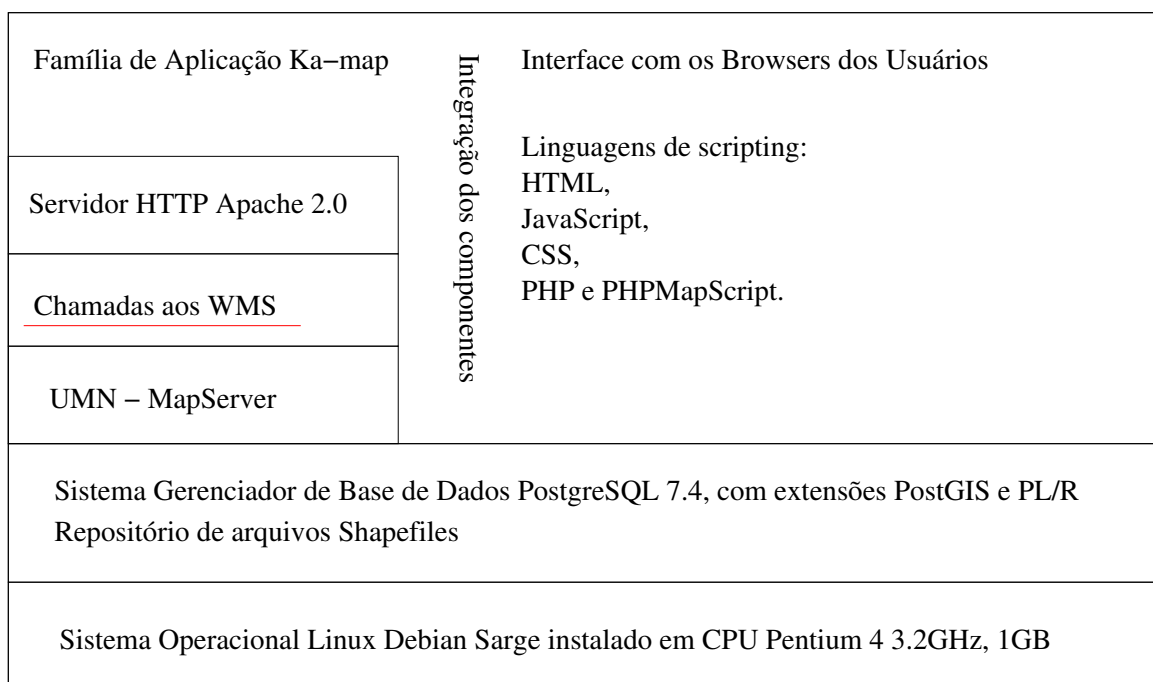


Figura 4.6: Arquitetura atual do protótipo GeoMedicina.

químicos são sobrepostas pelas camadas da base cartográfica com as fronteiras e com o nome dos municípios. Esta sobreposição de camadas aliada à classificação dos teores em uma escala que inicia pela cor cinza (menor teor do elemento), passa pelas cores azul, verde, amarela e laranja e termina com a cor vermelha (maior teor do elemento), permite identificar em quais municípios há uma maior concentração do elemento.

A integração de imagens satélite obtidas de serviços WMS provê uma visão que auxilia a compreensão dos mecanismos que provocam a concentração de determinado elemento em uma região, como por exemplo, uma represa ou um vale, ou a dispersão do mesmo através do curso dos rios e afluentes. As figuras 4.9 e 4.10 exemplificam esta visão.

A observação da figura 4.7 indica a concentração do elemento Bário na mesma região de trecho da bacia do Rio Paraná. A observação deste rio e seus afluentes pode auxiliar na localização da fonte do elemento bem como do provável destino deste material. O mapa da figura 4.10 tem como camada de fundo o tema Bário em águas. Neste caso, a camada superior constituída pela imagem satélite teve sua opacidade diminuída para que a camada inferior fosse sutilmente visualizada.

Outra utilidade de imagens satélite no sistema GeoMedicina é a identificação visual do uso

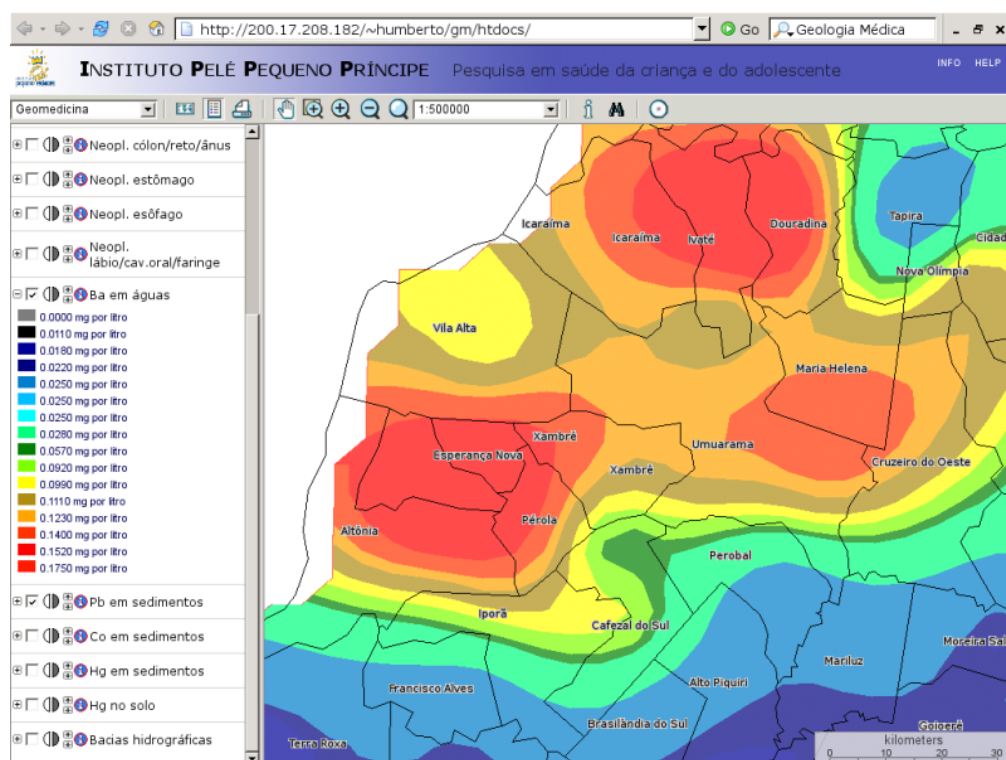


Figura 4.7: Mapa com a distribuição espacial do elemento Bário em águas. Fonte: MINEROPAR 1996.

de solo, como áreas de agricultura, áreas de florestas ou áreas urbanas. Esta visualização auxilia a avaliação da etiologia de determinada doença, se está associada a hábitos da população urbana ou rural. A figura 4.11 exemplifica esta identificação. É nítida a linha que separa a floresta (ao Sul) da área com plantações (ao Norte). Também é visível a área de represa (a Noroeste). A observação da figura 4.12 permite a identificação de municípios com área urbana predominante.

Em outro exemplo, a figura 4.13 ilustra a identificação de pontos onde há maior concentração do elemento Alumínio. O tamanho de fonte maior indica que o nível do elemento é maior nesta região em comparação com o nível detectado nas regiões onde o tamanho da fonte é menor.



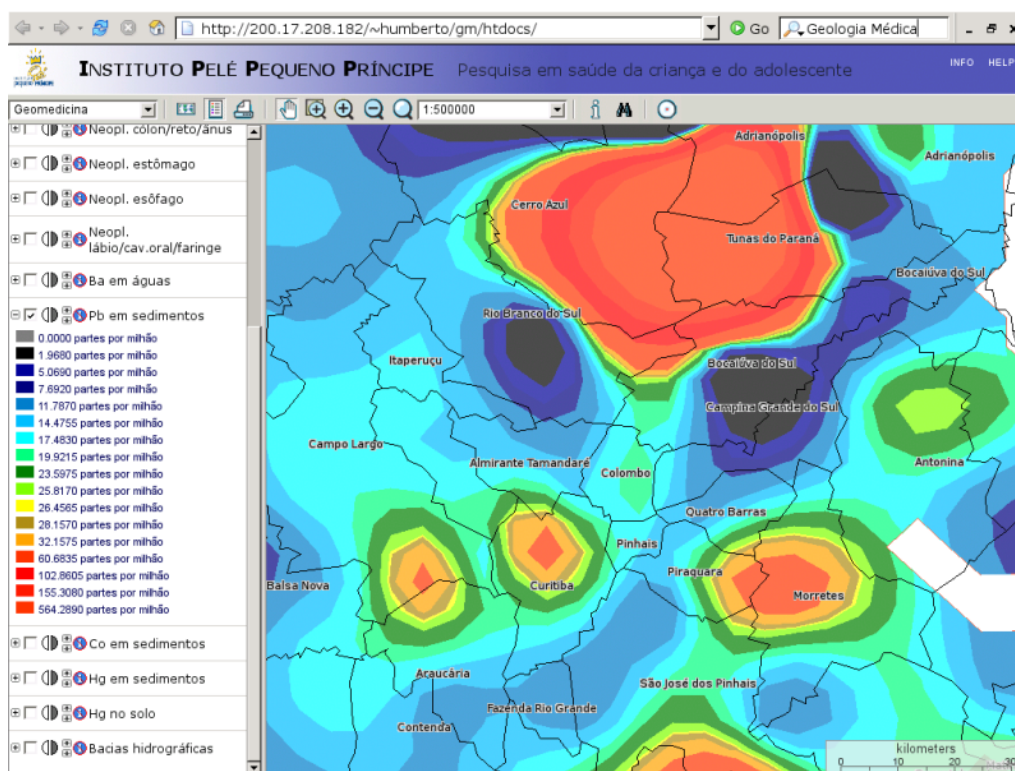


Figura 4.8: Mapa com a distribuição espacial do elemento Chumbo em sedimentos de fundo dos rios. Fonte: MINEROPAR 1996.

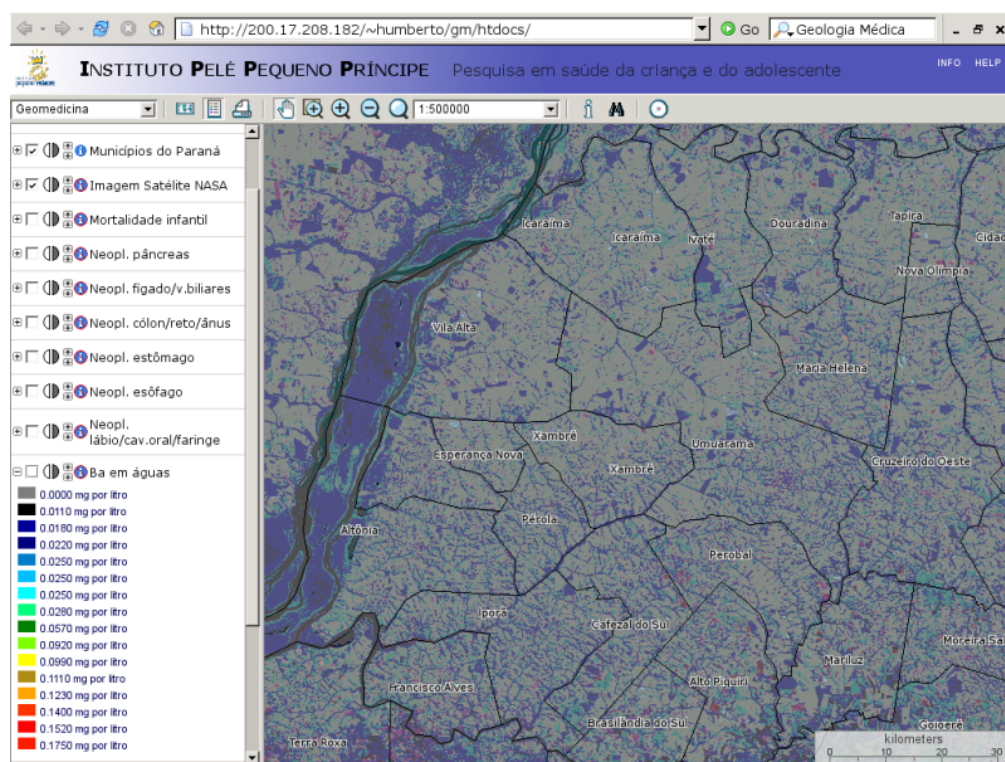


Figura 4.9: Mapa com imagem satélite de trecho da bacia do Rio Paraná. Fonte JPL-NASA 2004.

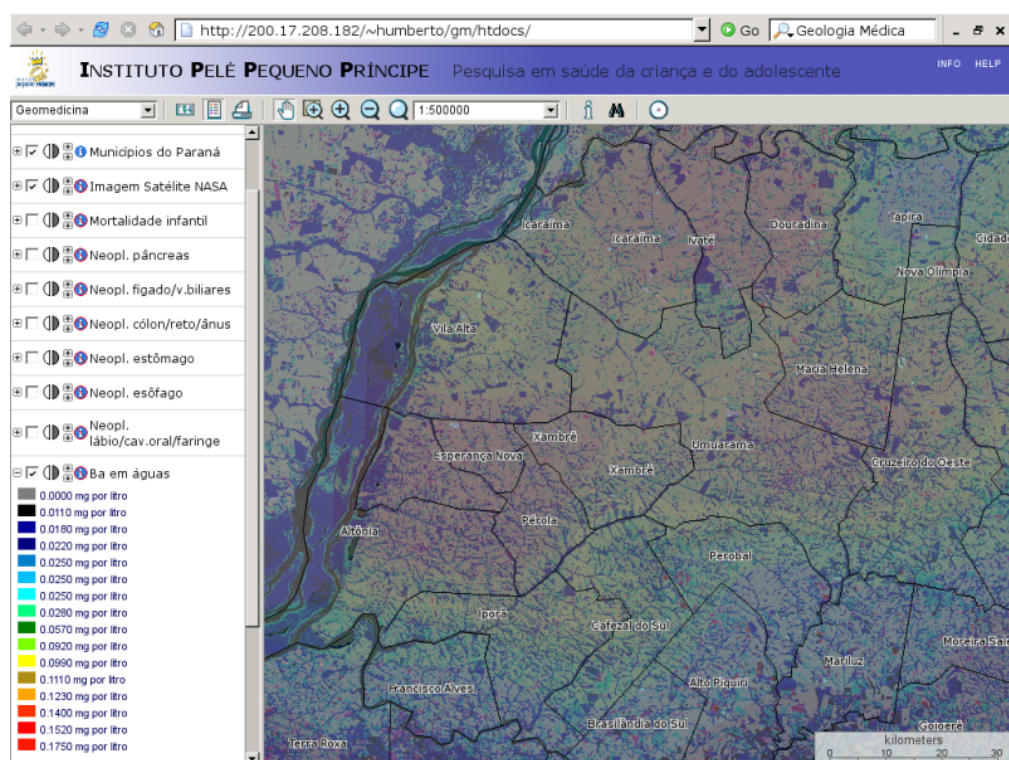


Figura 4.10: Mapa com imagem satélite transparente de trecho da bacia do Rio Paraná e com imagem de fundo teor do elemento Bário. Fonte JPL-NASA 2004 e MINEROPAR 1996.



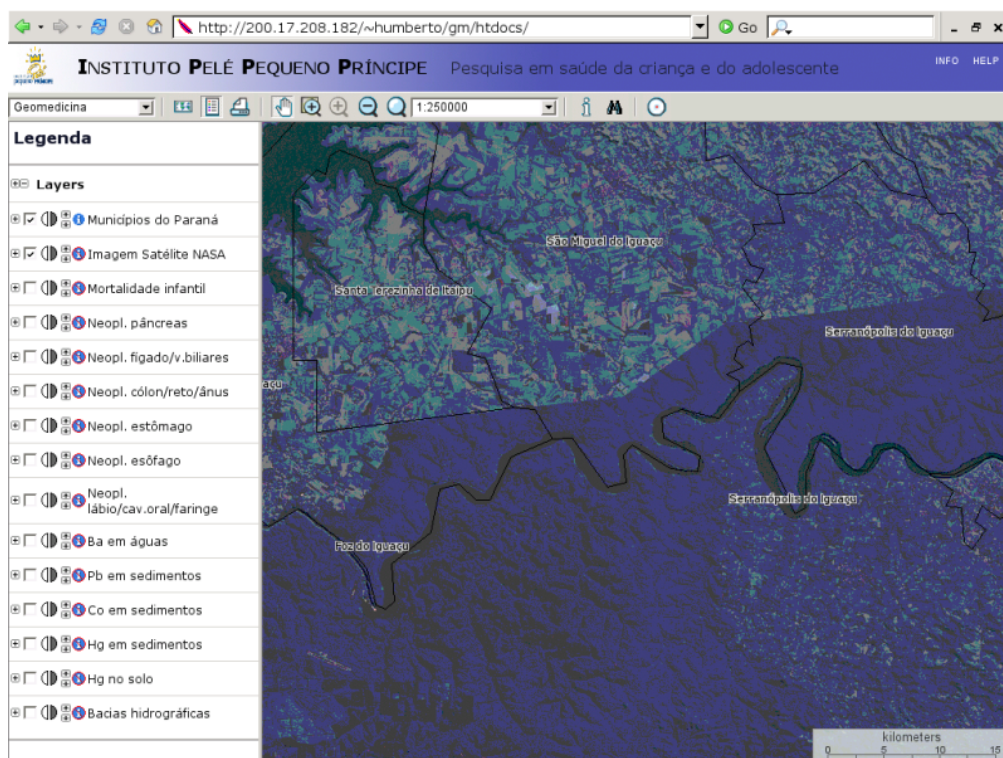


Figura 4.11: Mapa com imagem satélite identificando o uso do solo com área de floresta, área de agricultura e área de represa. Fonte: JPL-NASA 2004.

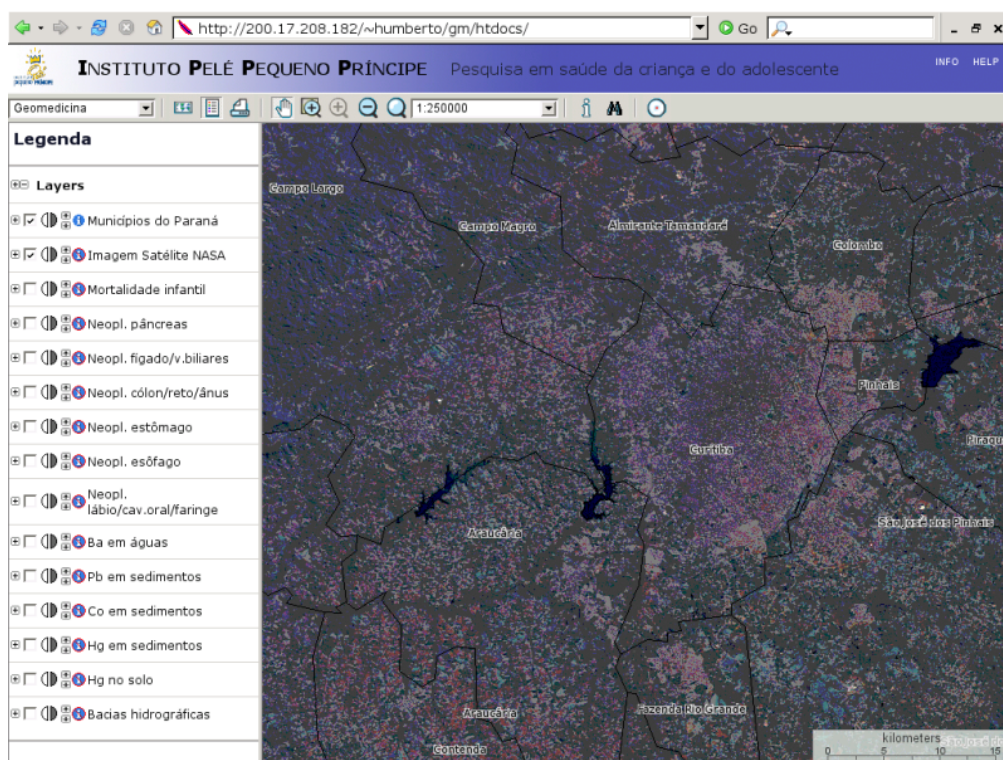


Figura 4.12: Mapa com imagem satélite identificando município com área urbana predominante. Fonte: JPL-NASA 2004.

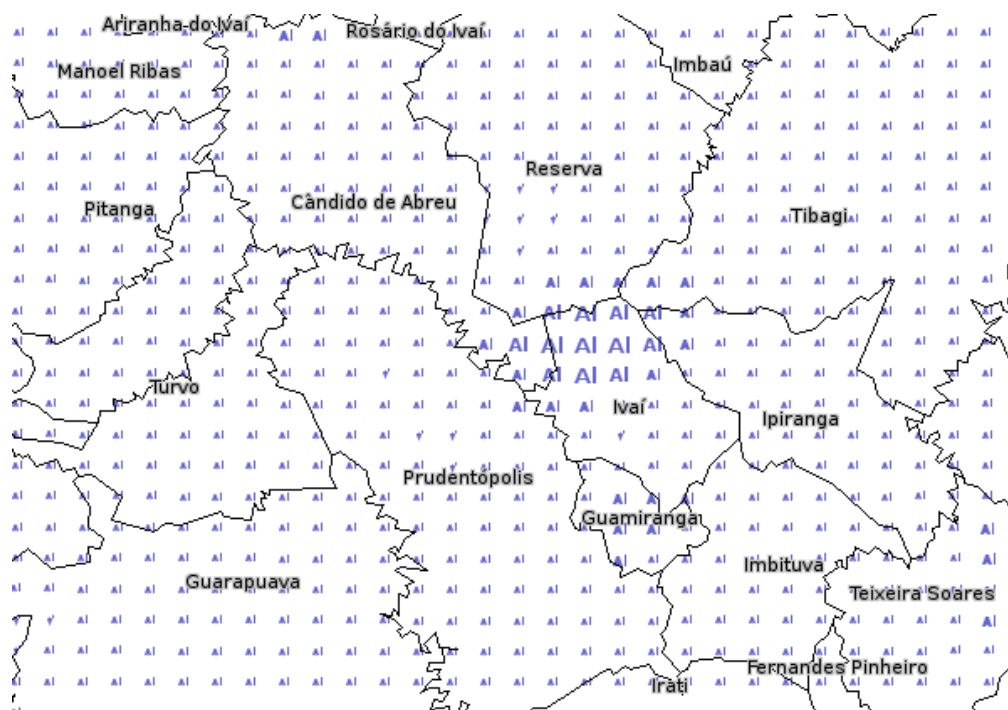


Figura 4.13: Mapa com a classificação da concentração de alumínio em uma região. Fonte: MINEROPAR 1996.

## 4.4 Avaliação do Protótipo

A versão atual do protótipo foi avaliada por usuários que responderam um questionário com acesso via *Web*, composto de 25 questões. Em cada questão o usuário expressou sua satisfação ao usar o protótipo, quantificando-a em uma escala de 1 a 10 ou N/A (não se aplica). O conteúdo do questionário é apresentado a seguir:

### Questionário para Avaliação do Protótipo GeoMedicina

Por favor, quantifique sua satisfação com o uso do protótipo GeoMedicina. A sua avaliação contribuirá para o aperfeiçoamento do protótipo.

- \* Tente responder a todas as questões.
- \* Se houver algum item julgado não aplicável, selecione: N/A
- \* Assegure-se de que o campo E-mail: esteja preenchido com o seu email.
- \* Se desejar, acrescente um comentário clicando sobre o ícone Acrescentar Comentário ao lado do item.
- \* Para enviar suas respostas, clique no botão Enviar Respostas, localizado abaixo das questões.

E-mail: [                      ]





	nunca há	o o o o o o o o o o	sempre há	o
20. Mensagens de ajuda na interface	não ajudam	o o o o o o o o o o	ajudam	o
21. Materiais suplementares, como manuais e tutoriais	confusos	o o o o o o o o o o	claros	o
CAPACIDADES DO PROTÓTIPO				
		1 2 3 4 5 6 7 8 9 10		N/A
22. Velocidade de navegação	muito lento	o o o o o o o o o o	rápido	o o
23. Conformidade com a realidade	inconforme	o o o o o o o o o o	conforme	o
24. O protótipo tende a ser	ruidoso	o o o o o o o o o o	silencioso	o
25. Correção de erros	difícil	o o o o o o o o o o	fácil	o
Liste os aspectos mais negativos:		Liste os aspectos mais positivos:		
1. [                      ]	1. [                      ]			
2. [                      ]	2. [                      ]			
3. [                      ]	3. [                      ]			

O gráfico da figura 4.14 apresenta os resultados da avaliação do protótipo. Quatorze usuários de diferentes áreas de atuação participaram da avaliação. A nota média do GeoMedicina para as 25 questões foi de 8,71, indicando que a navegação na interface foi agradável. Com base nas respostas apresentadas, pode-se observar que a maioria dos itens questionados recebeu uma avaliação boa ou ótima. As questões com melhor avaliação, média igual ou superior a 9,0 e nota mínima igual ou superior a 7,0, foram referentes à satisfação com o uso, a facilidade de leitura dos textos apresentados, a nitidez das imagens apresentadas, a conformidade com a realidade e a ausência de ruídos na interface.

Dentre os itens que receberam as notas mais baixas estão: a impressão geral de dificuldade ao usar o protótipo, a relação fraca entre a terminologia e as tarefas, a indicação confusa da tarefa a ser executada e a ausência de informação sobre o progresso de uma ação. Outros itens merecem a atenção para que seja melhorada a apresentação na interface do protótipo, entre eles, a adequação das informações apresentadas, a presença de telas de ajuda e a utilidade das mensagens de erro.

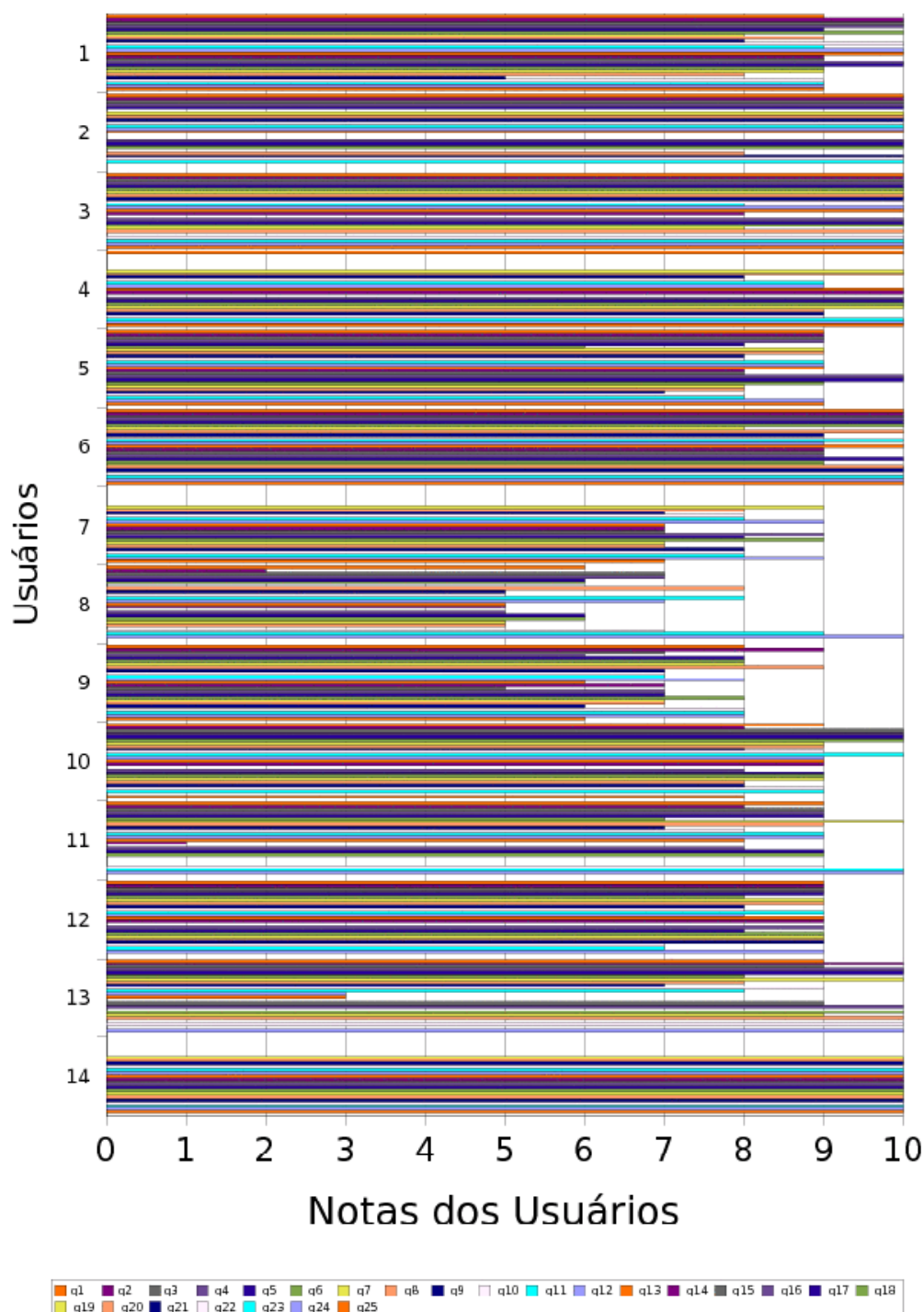


Figura 4.14: Gráfico com resultados das avaliações da satisfação dos usuários com o protótipo.

A tabela 4.1 apresenta as notas atribuídas por profissionais de diferentes áreas de atuação. As três primeiras linhas da tabela correspondem, nesta ordem, às notas atribuídas por profissionais das áreas de Cartografia, Farmácia e Física, as sete linhas seguintes correspondem às

notas atribuídas por profissionais da Informática e as quatro últimas linhas às notas atribuídas por profissionais da Medicina.

Notas dos 25 itens do questionário de satisfação do usuário																								
01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
9	8	9	9	9	7	10	9	7	8	9	9	8	1	NA	8	9	9	NA	NA	NA	9	10	10	NA
NA	NA	NA	NA	NA	NA	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
9	8	10	10	10	10	9	9	8	9	10	9	9	9	NA	8	9	9	9	8	8	9	9	NA	8
9	10	10	10	9	10	8	9	8	10	9	10	10	9	9	10	10	9	9	8	5	9	9	9	9
10	10	10	10	10	NA	10	10	10	10	10	10	10	NA	NA	10	10	10	NA	8	10	10	10	NA	NA
9	10	9	10	10	8	10	8	7	9	8	3	3	NA	9	10	NA	10	9	10	NA	10	NA	10	NA
NA	NA	NA	NA	NA	NA	9	8	7	8	8	9	7	7	7	9	8	9	7	7	8	8	8	8	7
10	10	10	10	10	10	8	10	9	9	10	9	10	9	9	9	10	9	9	10	10	10	10	10	10
9	9	9	9	8	6	9	9	8	8	9	9	9	8	8	10	10	9	8	8	7	8	8	9	9
10	10	10	10	10	10	10	10	10	10	8	10	10	8	NA	10	10	10	8	10	NA	10	10	10	10
6	2	7	7	6	6	NA	8	5	5	8	7	5	5	NA	5	6	6	5	5	NA	7	9	10	NA
10	NA	NA	NA	NA	NA	10	10	8	8	9	9	10	10	NA	10	10	10	10	9	9	9	10	10	10
8	9	7	6	8	8	8	9	7	7	7	8	6	7	5	7	7	8	7	7	6	8	8	8	6
9	9	9	9	9	8	9	9	8	8	9	8	9	9	NA	9	8	9	9	9	9	7	7	9	NA

Tabela 4.1: Tabela com os resultados do questionário de satisfação do usuário.

## CAPÍTULO 5

### DISCUSSÃO SOBRE O TRABALHO

Este capítulo apresenta uma reflexão sobre este trabalho, onde são discutidos o emprego do software livre, a abordagem de desenvolvimento de software empregada no projeto, a interface com o usuário e a arquitetura do sistema, a contribuição para a ação multidisciplinar e uma proposta de trabalhos futuros.

#### 5.1 Discussão sobre os Resultados

No GeoMedicina são empregados softwares livres, que, por serem de domínio público, não exigem o pagamento de licenças, o que reduz o custo do projeto. Nesta redução de custo deve-se considerar todo o software de base que seria comprado caso não fossem empregados softwares livres, como o Sistema Operacional, o Sistema Gerenciador de Bases de Dados e o Servidor HTTP.

Além disso, devido ao código fonte do software livre ser aberto à comunidade acadêmica, à indústria e a qualquer usuário de computador no mundo, a atualização de software, pelo seu aperfeiçoamento ou inclusão de novas funcionalidades, é bastante dinâmica. Este dinamismo é alcançado graças à contribuição de usuários, seja por meio de sugestões aos criadores e mantenedores destes softwares, seja por meio de sua participação na equipe de desenvolvimento. Atualmente, fundações e consórcios formados por organizações acadêmicas e empresariais dão suporte ao desenvolvimento destes softwares. Esta opção pelo uso do software livre foi motivada pelo sistema operacional Linux, cuja atualização de seus pacotes é acelerada com a contribuição de programadores do mundo todo, fato que contribui para que, por exemplo, falhas de segurança sejam identificadas e corrigidas rapidamente.

A visualização geográfica é apresentada em mapas digitais, com acesso via rede de computadores, em uma arquitetura cliente-servidor e em um ambiente interativo, onde o usuário seleciona os aspectos relevantes para a sua pesquisa, obtendo uma visualização que lhe apresenta

padrões que anteriormente estavam ocultos. O uso da Internet simplifica a disseminação e facilita o acesso à informação. É possível visualizar os mapas através de um computador, em qualquer local onde haja uma conexão com a rede, sem a necessidade de instalação de aplicativos específicos. Com o uso de um simples navegador, o usuário consegue interagir com o sistema.

No processo de software, empregou-se o método de desenvolvimento incremental em que, a partir de um protótipo inicial, evoluiu-se para versões aperfeiçoadas em um modelo espiral. Neste modelo o usuário final participa das avaliações e validações, proporcionando condições para que o sistema seja realmente útil e aceito pela organização para a qual ele foi desenvolvido. A escolha do protótipo inicial no contexto do software livre conduziu a abordagem do desenvolvimento com reuso, em que uma família de aplicações é reaproveitada e, com a devida consideração dos riscos, garante a eficiência e a confiança no sistema.

Essa família de aplicações tem como seu componente principal o *UMN-MapServer*, o qual integra softwares de computação gráfica e de processamento de imagens com softwares de leitura de dados espaciais para a produção de imagens digitais interpretadas pelos navegadores dos clientes.

O sistema é concebido segundo a arquitetura cliente-servidor, com acesso de multiusuários via rede de computadores. Os dados espaciais são modelados segundo os conceitos de expansão, obtida com a interoperação entre sistemas, onde ocorre o compartilhamento de dados produzidos pelo sistema local e de dados produzidos por serviços de mapas na Internet. Para atingir estas características são adotados na modelagem os padrões internacionais mais aceitos atualmente pela comunidade de usuários, que são os padrões consensuais propostos em conjunto pelo *Open Geospatial Consortium* e *International Organization for Standardization*. Esta abordagem garante um tempo de vida longo para o sistema, uma vez que o mesmo adquire independência da aplicação e a capacidade de expansão proporcionada pela interoperação com outros sistemas.

No aperfeiçoamento do protótipo são empregados conceitos de *design* de interfaces, considerando-se os princípios básicos de familiaridade, consistência e orientação ao usuário que, quando corretamente aplicados, aumentam a confiança do usuário no sistema. Um

levantamento criterioso de requisitos ajuda na adequação da interface do sistema às necessidades dos usuários. As interfaces do GeoMedicina são desenvolvidas com o emprego de linguagens como SQL, PHP, JavaScript, HTML e XML. Para uma correta apresentação são considerados conceitos de semiótica e visualização geográfica na principal saída do sistema, que são as imagens digitais de mapas.

Este sistema, na concepção de seus idealizadores, consiste em um grande potencial para a área da Ciência da Computação, uma vez que oferece várias linhas de pesquisa. Com a extensão de sua capacidade através do acesso a fenômenos naturais e sociais, advindos dos servidores de conteúdo georreferenciados que fornecem as *Coverages* citadas na seção 2.8, as quais consistem em imagens de satélite ou dados vetoriais associadas a informações de fenômenos, será possível desenvolver métodos de análise que contribuirão para novas descobertas na Geologia Médica ou no Meio Ambiente. Além disso, com o acesso às imagens espectrais produzidas pelos satélites, métodos de processamento digital de imagens poderão ser empregados para a identificação de padrões associados às reflexões do solo ou da água, os quais indicarão regiões onde os esforços de pesquisa poderão ser direcionados.

Este trabalho possibilitou a agregação de pesquisadores das mais diversas áreas em um trabalho colaborativo, favorecendo o fortalecimento das equipes multidisciplinares e fomentando novos projetos de pesquisa.

## 5.2 Trabalhos Futuros

A próxima evolução do protótipo sugerida é prover ao pesquisador a capacidade de usar o GeoMedicina como um caderno de anotações eletrônico, onde ele demarca as prováveis áreas de risco, construindo o seu mapa sobre as combinações possíveis de mapas geoquímicos e de doenças e identificando as áreas onde há uma suposta relação causa-efeito entre fatores do ambiente e doenças. Esta delimitação pode ser armazenada na base de dados espacial na forma vetorial e visualizada pelos demais pesquisadores para a corroboração da hipótese.

Outra evolução sugerida é a adição de capacidade estatística ao sistema com as finalidades de suavizar os indicadores de saúde representados em mapas de polígonos irregulares e de diversos tamanhos, e de identificar os relacionamentos importantes entre as variáveis do

ambiente e as doenças ocorridas nos lugares. Os relacionamentos podem ser calculados por meio dos modelos de regressão lineares citados na seção 2.6.

Na versão atual do GeoMedicina, os coeficientes de doenças (número de casos para cada 100 mil habitantes) são expressos em mapas coropléticos, compartimentados pelas fronteiras dos municípios; os teores dos elementos químicos são expressos na representação isoplética, a qual é apropriada ao fenômeno geoquímico. Ocorre que a representação coroplética assume que a incidência da doença está distribuída de forma uniforme na área do município, o que, na maioria das vezes, não corresponde à realidade, considerando-se a extensão dos municípios e as diferentes densidades das populações urbanas e rurais.

Outro tratamento estatístico desejado é a transformação da representação de área em ponto que poderia ser aplicada para fazer com que o valor da média da representação coroplética seja redistribuído em subdivisões da área, como os setores censitários, com uma ponderação pela população residente nestas subáreas. Esta idéia foi proposta por Goovaerts [42]. Nesta ponderação também poderia ser empregado o processamento digital de imagens para a identificação das subáreas de concentração urbana, visíveis nas imagens de satélite.

Com o emprego do protótipo pela comunidade científica, provavelmente haverá um aumento das variáveis de entrada do sistema e, conseqüentemente, do número de camadas selecionáveis pelo usuário. Uma nova interface com o usuário que constituirá a página inicial do sistema deverá ser desenvolvida para o controle das variáveis de entrada e do comportamento do sistema com o tratamento estatístico. Além disso, poderá ser empregado o estilo de linguagem de comandos caso, na evolução dos requisitos da interface, seja detectada a necessidade de flexibilizar as consultas espaço-temporais, permitindo ajustes finos nos dados de entrada da estatística aplicada para o estabelecimento das relações causa-efeito entre o ambiente e as doenças.

Também é necessário o aprimoramento da interface com o usuário, de forma que este possa construir o seu próprio mapa, inserindo pontos ou delimitando áreas onde os resultados da computação estatística e da visualização geográfica sugerirem associações entre os fatores ambientais e a incidência de doenças. Com esta possibilidade os usuários compartilharão suas hipóteses por meio dos mapas em um trabalho colaborativo, apresentado na seção 2.9. Neste



contexto, é desejável que pesquisadores possam reproduzir os processos executados por outros no sentido de validar e aumentar a confiança nos resultados obtidos. Isto poderá ser obtido com a implementação de um controle do fluxo e da proveniência dos dados conforme visto na seção 2.10.

## CAPÍTULO 6

### CONCLUSÕES

Este trabalho formulou as linhas mestres para a execução do projeto GeoMedicina, um projeto multidisciplinar pautado por disciplinas da Ciência da Computação como Engenharia de Software, Computação Gráfica, Sistemas de Bases de Dados, Interface Humano-Computador, Linguagens de Computação, mas que, por sua aplicação prática, foi permeado por conhecimentos de outras disciplinas como Geologia Médica, Estatística e Cartografia.

A seguir, discorre-se brevemente sobre a contribuição do projeto com relação aos objetivos específicos formulados.

A caracterização da distribuição espacial de mortalidade para doenças crônicas, por município do Estado do Paraná, foi obtida em mapas coropléticos com três classes: vermelha para os casos mais acentuados de incidência da doença em relação ao número de habitantes; amarela para os casos intermediários; e verde para os casos de menor incidência. É possível verificar no protótipo atual a distribuição espacial desta classificação para um conjunto das neoplasias constantes na CID-BR, a saber: pâncreas; fígado e vias biliares; cólon de intestino, reto e ânus; estômago; e lábio, cavidade oral e faringe. O protótipo atual comporta a adição de novas doenças.

A caracterização da distribuição espacial de elementos químicos foi representada por dois tipos de mapas. Os mapas de isolinhas, onde se observa uma classificação em dezesseis classes, representando a maior ou menor concentração do elemento químico. Neste tipo de mapa foram representados os elementos Bário (Ba), Chumbo (Pb), Cobalto (Co) e Mercúrio (Hg). Em um segundo tipo de mapa, a maior ou menor concentração do elemento é representada pelo tamanho da fonte referente ao elemento químico. Neste tipo de mapa foram representados os elementos Alumínio e Íon Nitrito (NO<sub>2</sub>).

A exibição das relações importantes entre variáveis químicas e de saúde é obtida com a sobreposição de pares de camadas elemento químico e doença, com o emprego do recurso de

transparência da camada superior.

A sugestão de hipóteses de relacionamento entre doenças e fatores ambientais contribui para a delimitação das supostas áreas de risco à saúde humana. Esta sugestão é enriquecida por recursos como a visualização de imagem satélite e a identificação do nome do município na área delimitada. Com a visualização da imagem satélite é possível identificar outros fatores do ambiente, como por exemplo, áreas urbanas, rurais, ou de floresta.

O GeoMedicina é uma ferramenta de exploração científica e de notificação das descobertas de áreas de risco à saúde humana. A exploração científica auxilia a pesquisa pelos agentes etiológicos das doenças crônicas, como as neoplasias, através da visualização geográfica: das supostas fontes contaminantes (agentes etiológicos), da incidência das doenças, e das hipóteses ou comprovações de relacionamentos causa-efeito entre os agentes etiológicos e as doenças. A notificação é feita pela identificação de áreas de risco à saúde humana pelos pesquisadores que podem encaminhá-las aos agentes da Vigilância em Saúde, responsáveis pelas ações sociais de informação e assistência à população atingida.

## REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ANA. Agência Nacional de Águas, Acesso em novembro de 2006. <http://www.ana.gov.br/>.
- [2] ANDERSON, B. A Comparison of ArcIMS to MapServer, Acesso em março de 2007. <http://mapserver.gis.umn.edu/community/conferences/MUM3/present/session10/msvsarcims/download>.
- [3] APACHE SOFTWARE FOUNDATION. The Apache HTTP Server Project, Acesso em novembro de 2006. <http://httpd.apache.org>.
- [4] ARCIMS. ArcIMS - Publish Maps, Data, and Metadata on the Web, Acesso em março de 2007. <http://www.esri.com/software/arcgis/arcims/>.
- [5] ARCSDE. ArcSDE - Advanced Spatial Data Server, Acesso em março de 2007. <http://www.esri.com/software/arcgis/arcsde/>.
- [6] ARCVIEW. Desktop GIS for Mapping, Data Integration, and Analysis, Acesso em outubro de 2006. <http://www.esri.com/software/arcview/>.
- [7] BAILEY, T. C. E GATRELL, A. C. *Interactive Spatial Data Analysis*. Harlow and Longman, Essex, Reino Unido e Nova York, NY, Estados Unidos, 1995.
- [8] BELL, B. S., HOSKINS, R. E., PICKLE, L. W. E WARTENBERG, D. Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public. *International Journal of Health Geographics* 5 (Novembro 2006), 49.
- [9] BERTIN, J. *Semiologie graphique. Les diagrammes - les reseaux - les cartes*. Gauthier - Villars, Paris, França, 1967.
- [10] BIVAND, R., ANSELIN, L., BERKE, O., BERNAT, A., CARVALHO, M., CHUN, Y., DORMANN, C., DRAY, S., HALBERSMA, R., LEWIN-KOH, N., ONO,

- H., PERES-NETO, P., TIEFELSDORF, M. E YU, D. SPDEP: Spatial dependence: weighting schemes, statistics and models R, Acesso em novembro de 2006. <http://cran.r-project.org/src/contrib/Descriptions/spdep.html>.
- [11] BOSCOE, F. P., WARD, M. H. E REYNOLDS, P. Current practices in spatial analysis of cancer data: data characteristics and data sources for geographic studies of cancer. *International Journal of Health Geographics* 3 (Dezembro 2004), 28.
- [12] CENTER FOR SPATIALLY INTEGRATED SOCIAL SCIENCE. Spatial Resources for the Social Sciences, Acesso em março de 2007. <http://www.csiss.org>.
- [13] CERVI, A., HEMSDORFF, H. H. M. E RIBEIRO, R. C. L. Tendência de Mortalidade por Doenças Neoplásicas em 10 Capitais Brasileiras, de 1980 a 2000. *Revista Brasileira de Epidemiologia* 8, 4 (Dezembro 2005), 407–418.
- [14] CGI. Common Gateway Interface CGI, Acesso em novembro de 2006. <http://hoohoo.ncsa.uiuc.edu/cgi/overview.html>.
- [15] CGIS. Canadian Geographic Information Systems, Acesso em fevereiro de 2007. [http://http://en.wikipedia.org/wiki/Canadian\\_Geographic\\_Information\\_Systems](http://http://en.wikipedia.org/wiki/Canadian_Geographic_Information_Systems).
- [16] CHANG, C.-L., CHANG, Y.-H., CHUANG, T.-R., DONG-PO DENG, A. W.-C. H. E HUANG, C.-H. The Web and Collaborative Geospatial Mapping. TR-IIS-06-011, Institute of Information Science, Academia Sinica, Taipei, Taiwan, Setembro 2006.
- [17] CHRISHMAN, N. R. History of Harvard Laboratory for Computer Graphics: a Poster Exhibit, Acesso em fevereiro de 2007. [http://isites.harvard.edu/fs/docs/icb.topic39008.files/History\\_LCG.pdf](http://isites.harvard.edu/fs/docs/icb.topic39008.files/History_LCG.pdf).
- [18] CIÊNCIA HOJE. Detecção precoce do câncer de supra-renal, Acesso em dezembro de 2006. <http://cienciahoje.uol.com.br/4183>.

- [19] CORRÊA, W. Análise Apache versus IIS, Acesso em março de 2007. <http://nadamais.wordpress.com/2006/06/23/analise-apache-vs-iis/>.
- [20] CÂMARA, G., SOUZA, R. C. M., FREITAS, U. M., GARRIDO, J. E II, F. M. SPRING: Integrating remote sensing and GIS by object-oriented data modeling. *Computers & Graphics* 20, 3 (Maio/Junho 1996), 395–403. <http://www.dpi.inpe.br/geopro/trabalhos/spring.pdf>.
- [21] DA SILVA RAMOS, C. *Visualização cartográfica e cartografia multimídia: conceitos e tecnologias*. Editora Unesp, São Paulo, SP, Brasil, 2005.
- [22] DATASUS. Departamento de Informática do SUS, Acesso em outubro de 2006. <http://w3.datasus.gov.br/datasus/datasus.php>.
- [23] DE BARROS SILVA, A. *Sistemas de Informações Geo-referenciadas: conceitos e fundamentos*. Editora Unicamp, Campinas, SP, Brasil, 2003.
- [24] DE FÁTIMA PINA, M. E SANTOS, S. M. *Conceitos Básicos de Sistemas de Informação Geográfica e Cartografia Aplicados à Saúde*. Organização Pan-Americana da Saúde, Brasília, DF, 2000.
- [25] DEBIAN. Debian GNU Linux: a free operating system (OS), Acesso em novembro de 2006. <http://www.debian.org>.
- [26] DERNIER, R. Using Regular Expressions in PostgreSQL, Acesso em março de 2007. <http://http://www.oreillynet.com/pub/a/databases/2006/02/02/postgresq-regexes.html?page=1>.
- [27] DEvesa, S. S., GRAUMAN, D. J., BLOT, W. J., PENNELLO, G. A., HOOVER, R. N. E FRAUMENI-JR, J. F. *Atlas of cancer mortality in the United States, 1950-94*. US Govt Print Off, Washington, DC, Estados Unidos, 1999.
- [28] DM SOLUTIONS GROUP. Ka-Map: An framework that is aimed at providing a javascript API for developing highly interactive web-mapping interfaces, Acesso em novembro de 2006. <http://ka-map.maptools.org>.

- [29] DOBSON, A. J. *An Introduction to Generalized Linear Models*, segunda ed. Chapman and Hall/CRC, Londres, 2001.
- [30] EMBRAPA. Embrapa - Monitoramento por Satélite, Acesso em novembro de 2006. <http://www.cnpm.embrapa.br/>.
- [31] ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE, INC. Introduction to Map Design, 1996. <http://www.esri.com/industries/k-12/download/docs/intrcart.pdf>.
- [32] ESRI. Environmental Systems Research Institute, Inc, Acesso em novembro de 2006. <http://www.esri.com/about.esri.html>.
- [33] FASTCGI. FastCGI an open extension to CGI that provides high performance for all Internet applications, Acesso em março de 2007. <http://www.fastcgi.com>.
- [34] GARRETT, J. J. AJAX: A New Approach to Web Applications, Acesso em fevereiro de 2007. <http://www.adaptivepath.com/publications/essays/archives/000385.php>.
- [35] GD. GD Graphics Library, Acesso em março de 2007. <http://www.boutell.com/gd/>.
- [36] GDDAL. Geospatial Data Abstraction Library - GDAL, Acesso em março de 2007. <http://www.gdal.org/>.
- [37] GEOMED. Paisagens Geoquímicas e Ambientais do Vale do Ribeira, Acesso em outubro de 2006. <http://www.ige.unicamp.br/geomed/>.
- [38] GEOMEDICINA. Áreas de Risco à Saúde Relacionadas com Fatores Ambientais, Constituição de uma Rede de Pesquisas e Implementação de um SIG. CV 2606, Acesso em dezembro de 2006. <http://www.softpar.com.br/UGF/site/projetos.php?idr=66&idp=72>.
- [39] GEOS. Geometry Engine Open Source - GEOS, Acesso em março de 2007. <http://geos.refractory.net/>.

- [40] GEOSERVER. GeoServer Open Source - connects your information to the Geospatial Web, Acesso em março de 2007. <http://docs.codehaus.org/display/GEOS/Home>.
- [41] GIULIO, G. M. D. Divulgação científica e comunicação de risco - um olhar sobre Adrianópolis, Vale do Ribeira. Dissertação de Mestrado, Universidade Estadual de Campinas, Campinas, SP, Brasil, 2006.
- [42] GOOVAERTS, P. Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging. *International Journal of Health Geographics* 5, 52 (Novembro 2006), 1-31.
- [43] IBGE. Instituto Brasileiro de Geografia e Estatística - Mapas Interativos, Acesso em novembro de 2006. <http://www.ibge.gov.br/mapas/>.
- [44] IJHG. International Journal of Health Geographics, Acesso em dezembro de 2006. <http://www.ij-healthgeographics.com/>.
- [45] JMETER. Apache JMeter - A Java application to load test functional behavior and measure performance, Acesso em março de 2007. <http://jakarta.apache.org/jmeter>.
- [46] JOLMA, A., AMES, D. P., HORNING, N., NETELER, M., RACICOT, A. E SUTTON, T. Free and Open Source Geospatial Tools for Environmental Modeling and Management, Acesso em março de 2007. [http://giscenter.isu.edu/training/pdf/Geotech\\_seminar/research\\_paper\\_dan\\_ames.pdf](http://giscenter.isu.edu/training/pdf/Geotech_seminar/research_paper_dan_ames.pdf).
- [47] KEPLER PROJECT. Kepler Project: a cross-project collaboration to develop open source tools for Scientific Workflows, Acesso em outubro de 2007. <http://kepler-project.org/Wiki.jsp?page=KeplerProject>.
- [48] KLIGERMAN, J. Estimates of cancer incidence and mortality in Brazil - 2000. *Revista Brasileira de Cancerologia* 46, 2 (Março 2000), 1.



- [49] KRAAK, M.-J. E ORMELING, F. *Cartography: visualization of spatial data*. Longman, Essex, Reino Unido, 1997.
- [50] LITCH, O. B. *Atlas Geoquímico do Estado do Paraná*. Minerais do Paraná, Curitiba, PR, Brasil, 2001.
- [51] MACHADO, A. C., BALICO, L. N. E IBAÑEZ, H. C. SVIM - Sistema de Visualização de Indicadores em Mapas, 2004. Trabalho de Conclusão de Curso Bacharelado em Ciência da Computação.
- [52] MAPGUIDE. MapGuide Open Source Project Home, Acesso em março de 2007. <http://mapguide.osgeo.org/>.
- [53] MAPSCRIPT. MapScript API Reference, Acesso em novembro de 2006. <http://mapserver.gis.umn.edu/docs/reference/mapscript>.
- [54] MCKENNA, J. Mapfile Reference, Acesso em novembro de 2006. <http://mapserver.gis.umn.edu/docs/reference/mapfile>.
- [55] MIME. MIME Media Types, Acesso em março de 2007. <http://www.iana.org/assignments/media-types/>.
- [56] MINEROPAR. Localização das células secundárias do Global Geochemical Reference Network, Acesso em novembro de 2006. <http://www.mineropar.pr.gov.br/modules/conteudo/conteudo.php?conteudo=6>.
- [57] MINISTÉRIO DE MEIO AMBIENTE. Web Service Client, Acesso em março de 2007. <http://mapas.mma.gov.br/i3geo/wsclient.htm>.
- [58] MITCHELL, T. *Web Mapping Illustrated*. O'Reilly & Associates, Sebastopol, CA, Estados Unidos, 2005.
- [59] MMAOGC. Ministério do Meio Ambiente - Web Services OGC, Acesso em novembro de 2006. <http://mapas.mma.gov.br/i3geo/menutemas/servicoswms.xml>.

- [60] MORAES, R. P., FIGUEIREDO, B. R. E LAFON, J.-M. Pb-Isotopic tracing of metal-pollution sources in the Ribeira Valley, southeastern Brazil. *Terrae* 1, 1 (2004), 26–33.
- [61] MORIMOTO, C. E. *Redes e Servidores Linux: Guia Prático*. Sul Editores, Porto Alegre, RS, Brasil, 2006.
- [62] NASA. NASA Jet Propulsion Laboratory California Institute of Technology, Acesso em novembro de 2006. <http://onearth.jpl.nasa.gov>.
- [63] NATIONAL CANCER INSTITUTE. Atlas of Cancer Mortality, Acesso em setembro de 2006. <http://cancercontrolplanet.cancer.gov/atlas/>.
- [64] NAVATHE, S. B. E ELMASRI, R. *Fundamentals of Database Systems*. Addison Wesley, Boston, MA, Estados Unidos, 2003.
- [65] NOAA. NOAA Satellite and Information Service, Acesso em novembro de 2006. <http://www.ngdc.noaa.gov/maps/>.
- [66] OASIS UDDI SPEC TC. UDDI Version 3.0.2, Acesso em março de 2007. [http://uddi.org/pubs/uddi\\_v3.htm](http://uddi.org/pubs/uddi_v3.htm).
- [67] OGC. Geographic Markup Language, Acesso em novembro de 2006. <http://www.opengeospatial.org/standards/gml>.
- [68] OGC. OGC Simple Feature Specification for SQL, Acesso em novembro de 2006. <http://www.opengeospatial.org/standards/sfs>.
- [69] OGC. OGC Web Coverage Service (WCS) Implementation Specification, Acesso em novembro de 2006. <http://www.opengeospatial.org/standards/wcs>.
- [70] OGC. OGC Web Feature Service Implementation Specification, Acesso em novembro de 2006. <http://www.opengeospatial.org/standards/wfs>.
- [71] OGC. OGC Web Map Service Implementation Specification, Acesso em novembro de 2006. <http://www.opengeospatial.org/standards/wms>.

- [72] OGC. Open Geospatial Consortium, Acesso em novembro de 2006. <http://www.opengeospatial.org/>.
- [73] OGR. OGR Simple Features Library, Acesso em março de 2007. <http://www.gdal.org/ogr/>.
- [74] O'REILLY XML.COM. Build AJAX-Based Web Maps Using ka-Map, Acesso em novembro de 2006. <http://www.xml.com/pub/a/2005/08/10/ka-map.html>.
- [75] PCI. Web Map Services, Acesso em fevereiro de 2007. <http://www.pcigeomatics.com/>.
- [76] PDFLIB. PDFlib, Acesso em março de 2007. <http://www.pdflib.com/>.
- [77] PETER BUNEMAN AND SANJEEV KHANNA AND WANG-CHIEW TAN. Why and Where: A Characterization of Data Provenance. In *Database Theory - ICDT 2001, 8th International Conference* (London, UK, 2001), J. V. den Bussche and V. Vianu, Eds., Springer, pp. 316–330. <http://db.cis.upenn.edu/DL/whywhere.pdf>.
- [78] PHP. PHP: Hypertext Preprocessor, Acesso em novembro de 2006. <http://www.php.net>.
- [79] PHP MAPSCRIPT. PHP Mapscript Class Diagram, Acesso em novembro de 2006. [http://www.veremes.com/download/php-mapscript\\_class\\_diagram-4.4\\_int.pdf](http://www.veremes.com/download/php-mapscript_class_diagram-4.4_int.pdf).
- [80] PHP MAPSCRIPT. PHP Mapscript dinamic loadable module, Acesso em novembro de 2006. [http://www.maptools.org/php\\_mapscript](http://www.maptools.org/php_mapscript).
- [81] PIANOVSKI, M. A., MALUF, E. M., DE CARVALHO, D. S., RIBEIRO, R. C., RODRIGUEZ-GALINDO, C., BOFFETTA, P., ZANCANELLA, P. E FIGUEIREDO, B. C. Mortality rate of adrenocortical tumors in children under 15 years of age in Curitiba, Brazil. *Pediatric Blood & Cancer* 47 (Setembro 2005), 56–60.
- [82] PLR. R Procedural Language to the PostgreSQL, Acesso em novembro de 2006. <http://www.jeoconway.com/plr>.

- [83] POSTGIS. PostGIS Extension to the PostgreSQL, Acesso em novembro de 2006. <http://www.refractions.net/postgis>.
- [84] POSTGRESQL. PostgreSQL Database Management System, Acesso em novembro de 2006. <http://www.postgresql.org>.
- [85] PRESSMAN, R. S. *Engenharia de Software*. Makron Books, São Paulo, 1995. Tradução de José Carlos Barbosa dos Santos e revisão técnica de Paulo Cesar Masiero e José Carlos Maldonado and Rosely Sanches.
- [86] PROJ4. PROJ4 - Cartographic Projections Library, Acesso em março de 2007. <http://proj.maptools.org/>.
- [87] R DEVELOPMENT CORE TEAM. *The R Project for Statistical Computing*. Viena, Viena, Austria, 2006. <http://www.R-project.org>.
- [88] RAMSEY, P. Introduction to PostGIS, Acesso em março de 2007. <http://www.refractions.net/postgis>.
- [89] RIBEIRO-JR, P. J. E DIGGLE, P. geoR: a package for geostatistical analysis. *R-News* 1, 2 (Junho 2001), 15–18.
- [90] RIGAUX, P., SCHOLL, M. E VOISARD, A. *Spatial Databases With Application to GIS*. Morgan Kaufmann, San Francisco, CA, Estados Unidos, 2002.
- [91] SAUDAVEL. Sistema de Apoio Unificado para Detecção e Acompanhamento em Vigilância Epidemiológica, Acesso em outubro de 2006. <http://www.dpi.inpe.br/saudavel/>.
- [92] SAYAR, A., PIERCE, M. E FOX, G. Integrating AJAX Approach into GIS Visualization Web Services. *IEEE Computer Society Proceedings of the Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services* (Fevereiro 2006), 169.

- [93] SELINUS, O., ALLOWAY, B., CENTENO, J. A., FINKELMAN, R. B., FULGE, R., LINDH, U. E SMEDLEY, P. *Essentials of Medical Geology: impacts of the natural environment on public health*. Elsevier Academic Press, San Diego, CA, Estados Unidos, 2005.
- [94] SESA. Secretaria de Estado da Saúde do Paraná, Acesso em outubro de 2006. <http://www.saude.pr.gov.br/>.
- [95] SHAPEFILE. ESRI Shapefile Technical Description, Acesso em novembro de 2006. <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>.
- [96] SHAWN BOWERS AND TIMOTHY MCPHILLIPS AND BERTRAM LUDASCHER AND SHIRLEY COHEN AND SUSAN B. DAVIDSON. A Model for User-Oriented Data Provenance in Pipelined Scientific Workflows. In *International Provenance and Annotation Workshop (IPAW'06)* (Chicago, Illinois, USA, 2006), L. Moreau and I. T. Foster, Eds., vol. 4145 of *LNCS*, Springer, pp. 133–147. [http://www.ipaw.info/ipaw06/proceedings/CameraReady\\_s5\\_3.pdf](http://www.ipaw.info/ipaw06/proceedings/CameraReady_s5_3.pdf).
- [97] SHNEIDERMAN, B. *Designing the User Interface*, terceira ed. Addison Wesley, Reading, MA, Estados Unidos, 2003.
- [98] SKYLAB MOBILESYSTEMS LTD. Latest OGC compliant WMS Servers, Acesso em novembro de 2006. [http://www.skylab-mobilesystems.com/en/wms\\_serverlist.html](http://www.skylab-mobilesystems.com/en/wms_serverlist.html).
- [99] SLOCUM, T. A. *Thematic Cartography and visualization*. Prentice Hall, Upper Saddle River, NJ, Estados Unidos, 1998.
- [100] SNELL, J. Call SOAP Web services with Ajax, Part 1: Build the Web services client. *DeveloperWorks IBM's resource for developers* (Outubro 2005).
- [101] SOARES, W. *AJAX (Asynchronous JavaScript and XML) Guia Prático Para Windows*. Erica, São Paulo, 2006.

- [102] SOMMERVILLE, I. *Engenharia de Software*. Addison Wesley, São Paulo, 2003. Tradução de André Maurício de Andrade Ribeiro e revisão técnica de Keechi Hiramã.
- [103] SVS. Secretaria de Vigilância em Saúde, Acesso em novembro de 2006. <http://portal.saude.gov.br/portal/svs/default.cfm>.
- [104] TC211. International Organization for Standards Technical Committee 211, Acesso em novembro de 2006. <http://www.isotc211.org/>.
- [105] TERRAVIEW. Aplicativo Visualizador da Dados Geográficos, Acesso em outubro de 2006. <http://www.dpi.inpe.br/terraview/>.
- [106] THE NETCRAFT WEB SERVER SURVEY. Web Server Count by Active Sites, Acesso em março de 2007. <http://survey.netcraft.com/index-200007.html#active>.
- [107] UC BERKELEY. GiST - Generalized Search Tree, Acesso em março de 2007. <http://gist.cs.berkeley.edu>.
- [108] UNIVERSITY OF MINNESOTA AND NASA TERRASIP PROJECT. UMN-MapServer: a open source development environment for building spatially-enabled Internet applications, Acesso em novembro de 2006. <http://ms.gis.umn.edu>.
- [109] W3C ARCHITECTURE DOMAIN. Web Services Description Language (WSDL) 1.1, Acesso em março de 2007. <http://www.w3.org/TR/wsdl>.
- [110] W3C ARCHITECTURE DOMAIN. XML Schema, Acesso em março de 2007. <http://www.w3.org/XML/Schema>.
- [111] WORLD WIDE WEB CONSORTIUM - W3C. Cascading Style Sheets, Acesso em março de 2007. <http://www.w3.org/Style/CSS/>.
- [112] WORLD WIDE WEB CONSORTIUM - W3C. Cascading Style Sheets, level 2 CSS2 Specification, Acesso em março de 2007. <http://www.w3.org/TR/REC-CSS2/>.
- [113] WORLD WIDE WEB CONSORTIUM - W3C. Document Object Model - DOM, Acesso em março de 2007. <http://www.w3.org/DOM/>.

- [114] WÜNSCH-FILHO, V. E ZAGO, M. A. Modern cancer epidemiological research: genetic polymorphisms and environment. *Revista Saúde Pública* 39, 3 (Junho 2005), 490–497.

## APÊNDICE A

### INSTALAÇÃO E CONFIGURAÇÃO DOS PACOTES DEBIAN

Neste apêndice são apresentados os procedimentos para instalação e configuração dos pacotes de software livre a partir do repositório Debian, bem como a configuração de aplicativos instalados para atender às necessidades peculiares do sistema. Esta instalação foi realizada em 2007 quando a versão estável do Debian era a Etch. Optou-se pela versão estável por sua robustez e segurança.

#### A.1 Instalação dos Pacotes Debian

Os pacotes são obtidos do repositório oficial e estável da distribuição Debian. No primeiro passo da instalação é feita a formatação do disco rígido, na qual o usuário escolhe o sistema de arquivos. Nesta instalação, optou-se pelo sistema EXT3. Em seguida é instalado o sistema básico com o reconhecimento automático do hardware e o carregamento dos respectivos *drivers*. Caso haja alguma falha de reconhecimento será necessário informar algumas características do hardware, por exemplo, o modelo da placa de vídeo ou da placa de rede. É recomendado ter estas informações em mãos, bem como o endereço IP e a máscara de rede que são solicitados no final desta etapa.

Em seguida, o computador é reiniciado com o sistema básico. A partir deste ponto, o usuário seleciona um conjunto de pacotes que deseja instalar, sendo possível instalá-los via Internet, caso a configuração da rede tenha sido bem sucedida. Nesta seleção, optou-se por um conjunto mínimo de pacotes que constituem um servidor HTTP. Feita a instalação do sistema básico e do servidor HTTP inicia-se a instalação dos pacotes que compõem o servidor de mapas. Esta instalação é feita via comando “apt-get” em uma *shell* do sistema operacional instalado.

O comando “apt-get install” requer alguns preparativos. É necessário atualizar uma lista com as datas de construção dos pacotes disponíveis no repositório antes de iniciar. Para



isso, executa-se o comando “apt-get update” que obtém a lista e compara-a com uma lista dos pacotes já instalados no computador, verificando os candidatos à atualização, ou seja, aqueles pacotes instalados e cuja data de construção é anterior a data do mesmo pacote na lista atualizada. Em seguida, faz-se a atualização dos candidatos com o comando “apt-get upgrade”. As atualizações são obtidas, descompactadas e instaladas no computador de forma automática. A vantagem dos comandos APT é o controle das dependências que instala sob demanda pacotes essenciais. Por isso, é natural, nas primeiras instalações, serem acrescentados à lista digitada pelo usuário pacotes que devem ser aceitos pelo usuário.

Caso o usuário tenha necessidade de versões de software mais recentes ainda não constantes nos repositórios oficiais Debian, alguns softwares podem ser instalados a partir de configuração e compilação dos fontes. Para isso, é necessário que bibliotecas requeridas estejam instaladas no servidor. Em alguns casos, há uma interrupção na compilação devido à falta de alguma biblioteca acompanhada de uma mensagem que identifica o *nome* da biblioteca ausente. Neste caso, pode-se usar o comando “apt-cache search *nome*” cujo resultado indicará o nome completo da biblioteca.

Segue uma lista dos comandos APT para instalação dos principais pacotes que constituem o servidor de mapas.

- Pacotes para processamento, leitura de formatos dados espaciais e conversão de sistemas de projeção: apt-get install proj libgdal1-1.3.2-dev gdal-bin.
- Pacotes para processamento, desenho da geometria e processamento das imagens digitais: apt-get install libpng12-dev libxpm-dev libgd2-xpm-dev libjpeg62-dev libfreetype6-dev zlib1g-dev libcfitsio3 giflib3g-dev php5-gd
- Pacotes para processamento de protocolos: apt-get install php5-dev libapache2-mod-php5 libiconv-hook-dev libxerces27-dev libjasper-1.701-dev libcurl3-dev netcdfg-dev
- Pacotes para gerenciamento e análise de dados espaciais: apt-get install postgresql postgresql-client postgresql-contrib postgresql-dev postgresql-8.1-plr r-base r-base-core r-cran-xml r-doc-info r-mathlib libgeos-dev libpq4 libpq-dev libapache2-mod-auth-pgsql php5-pgsql postgis postgresql-8.1-postgis libdbd-pg-perl libpg-perl

- Pacotes para serviço de mapas: `apt-get install mapserver-bin perl-mapscript python-mapscript php5-mapscript cgi-mapserver mapserver-doc libdbi-perl`

## A.2 Configuração do Servidor Web e Implantação do Framework

Recomenda-se para os testes iniciais, habilitar a publicação de conteúdo para Internet no diretório `public_html` do usuário com o comando `"a2enmod userdir"`, digitado em *shell* Linux como superusuário. Depois de habilitar o módulo, o arquivo `"/etc/apache2/sites-available/default"` deve ser editado para a inclusão das seguintes linhas:

```
<IfModule mod\_userdir.c>
    UserDir public\_html
</IfModule>
```

Após esta configuração, reinicia-se o serviço com o comando `"/etc/init.d/apache2 force-reload"`. Cria-se o diretório com o comando `"mkdir public_html"` e posiciona-se neste com o comando `"cd public_html"`.

Em seguida, faz-se o download do Ka-map e de um conjunto de mapas usados para demonstração, o Gmap, por meio dos comandos `"wget http://ka-map.maptools.org/dl/ka-map-1.0-20070205.tar.gz"` e `"wget http://dl.maptools.org/dl/gmap-ms46.tar.gz"`.

Descompacta-se os pacotes adquiridos no diretório `"public_html"` com os comandos `"tar -xvzf ka-map-1.0-20070205.tar.gz"` e `"tar -xvzf gmap-ms46.tar.gz"` e modifica-se os nomes dos diretórios criados na descompactação com os comandos `"mv ka-map-1.0-20070205 kamap"` e `"mv gmap-ms46 gmap"`.

Cria-se uma cópia do arquivo `"config.dist.php"` localizado em `"public_html/kamap/include"` com o nome `"config.php"` por meio dos comandos `"cd kamap/include"` e `"cp config.dist.php config.php"`. Edita-se este arquivo, procurando-se pelo padrão `"php_mapscript"`, o qual deve coincidir com o nome encontrado como resultado dos comandos `"updatedb"` e `"locate php_mapscript | grep so"`.

Agora, o servidor de mapas instalado e o *framework* Ka-map estão prontos para o teste. Inicia-se o navegador, apontando-o para `http://host/usuario/kamap/htdocs` e verifica-se a abertura da interface e a apresentação do mapa do Canadá na área de visualização geográfica.

## APÊNDICE B

### CRIAÇÃO DA BASE DE DADOS

A criação da base de dados do sistema GeoMedicina e a inserção de registros nesta base é codificada nas linguagens SQL e PHP, em uma seqüência apresentada nas próximas seções.

#### B.1 Criação da Base de Dados no PostGIS

A seguir é apresentada uma seqüência de comandos executados em *shell* para a criação e a configuração de uma base de dados espacial, incluindo comandos para o carregamento da linguagem PL/PGSQL e das definições de objetos PostGIS e para a criação das tabelas geometria ("geometry\_columns") e sistema de referência espacial ("spatial\_ref\_sys"). Para facilitar o processo de criação de outras bases de dados espaciais é criado o "templateGIS", o qual é usado como esqueleto na criação das mesmas.

```
su -
su - postgres
createlang plpgsql templateGIS
psql -d templateGIS -f "/usr/share/postgresql/contrib/lwpostgis.sql"
psql -d templateGIS -f "/usr/share/postgresql/contrib/spatial_ref_sys.sql"
psql templateGIS
geomedicina=# alter table spatial_ref_sys owner to humberto;
geomedicina=# alter table geometry_columns owner to humberto;
createdb -O humberto -T templateGIS geomedicina
```

A tabela geometria mostrada na estrutura a seguir é criada pelo *script* "lwpostgis.sql" e contém metadados referentes às tabelas de dados espaciais.

```
geomedicina=> \d geometry_columns
          Tabela "public.geometry_columns"
   Coluna          |          Tipo          | Modificadores
-----|-----|-----
f_table_catalog    | character varying(256) | not null
f_table_schema     | character varying(256) | not null
f_table_name       | character varying(256) | not null
f_geometry_column  | character varying(256) | not null
coord_dimension    | integer                | not null
srid               | integer                | not null
type               | character varying(30)  | not null
attrelid           | oid                   |
varattnum          | integer               |
stats              | histogram2d            |
Índices:
    "geometry_columns_pk" chave primária,
    btree (f_table_catalog, f_table_schema, f_table_name, f_geometry_column)
```

A tabela sistema de referência espacial é criada pelo *spatial\_ref\_sys* e contém a definição dos códigos *spatial reference ID* (SRID) mostradas na estrutura a seguir:

```
geomedicina=> \d spatial_ref_sys;
Tabela "public.spatial_ref_sys"
Coluna      | Tipo      | Modificadores
-----|-----|-----
srid         | integer   | not null
auth_name    | character varying(256)
auth_srid    | integer
srtext       | character varying(2048)
proj4text    | character varying(2048)
Índices:
    "spatial_ref_sys_pkey" chave primária, btree (srid)
```

Os valores dos campos “srtext” referentes aos códigos SRID 4261 e 29182, empregados na base de dados GeoMedicina, ambos definidos pela autoridade *European Petroleum Survey Group* (EPSG), são apresentados a seguir:

```
SRID 4261 - GEOGCS "SAD69"
DATUM "South_American_Datum_1969"
SPHEROID "GRS 1967", 6378160 298.247167427, AUTHORITY "EPSG" "7036", AUTHORITY "EPSG" "6291"
PRIMEM "Greenwich", 0, AUTHORITY "EPSG" "8901"
UNIT "degree", 0.0174532925199433, AUTHORITY "EPSG" "9108", AUTHORITY "EPSG" "4291"
SRID 29182 - PROJCS "SAD69 / UTM zone 22S"- GEOGCS "SAD69"
DATUM "South_American_Datum_1969"
SPHEROID "GRS 1967", 6378160 298.247167427, AUTHORITY "EPSG", "7036", AUTHORITY "EPSG" "6291"
PRIMEM "Greenwich", 0, AUTHORITY "EPSG" "8901"
UNIT "degree", 0.0174532925199433, AUTHORITY "EPSG" "9108", AUTHORITY "EPSG", "4291"
PROJECTION "Transverse_Mercator"
PARAMETER "latitude_of_origin" 0
PARAMETER "central_meridian" -51
PARAMETER "scale_factor" 0.9996
PARAMETER "false_easting" 500000
PARAMETER "false_northing" 10000000
UNIT "metre" 1, AUTHORITY "EPSG" "9001", AUTHORITY "EPSG" "29182"
```

## B.2 Criação de Tabelas a partir de ShapeFile

Na instalação do PostGIS, adquire-se uma ferramenta para a conversão de arquivos *ShapeFile* em uma tabela da base de dados PostGIS. O procedimento consiste em executar o conversor “shp2pgsql” que cria uma tabela com objetos geográficos e atributos equivalentes aos do arquivo original. Segue o comando usado para converter um *ShapeFile* obtido do IBGE.

```
shp2pgsql -s 4291 -c -D parana.shp parana | psql geomedicina
```

## B.3 Criação de Tabelas a partir de Arquivo DBF

Uma importação de dados para a base foi realizada com uma compilação de dados sobre mortalidade por doenças em um arquivo no formato DBF cujos campos de interesse são: o número seqüencial do óbito no Estado naquele ano (NUMERODO), o município de residência do paciente (CODMUNRES) e a causa básica da morte segundo a classificação da CID-10 (CAUSABAS). Outros dados étnicos e sócio-econômicos como idade, sexo, raça e escolaridade

do paciente constam no DBF mas não foram considerados nesta etapa. Foram fornecidos DBFs dos anos de 1996 ao ano de 2004.

O procedimento para importação dos arquivos DBF é realizado com um conversor, o “dbf2pg”, que cria uma tabela em base de dados PostgreSQL com todos os registros do DBF original, inserindo-os em campos do tipo caracter. A estrutura da tabela *dopr2004* da base de dados *geomedicina* mostra os campos criados com o “dbf2pg”. Segue o comando usado para a conversão:

```
dbf2pg -d geomedicina -t dopr2004 DOPR2004.DBF
```

A seguir são mostrados o esquema da base de dados e a estrutura da tabela importada do arquivo DBF.

#### - Esquema da base de dados *geomedicina*

```
humberto@geomed:~$ psql geomedicina
```

```
geomedicina=> \d
```

Esquema	Nome	Tipo	Dono
public	c00c14	tabela	postgres
public	c15	tabela	postgres
public	c16	tabela	postgres
public	c18c21	tabela	postgres
public	c22	tabela	postgres
public	c25	tabela	postgres
public	domun1996	tabela	postgres
public	domun2004	tabela	postgres
public	dopr1996	tabela	postgres
public	dopr1997	tabela	postgres
public	dopr1998	tabela	postgres
public	dopr1999	tabela	postgres
public	dopr2000	tabela	postgres
public	dopr2001	tabela	postgres
public	dopr2002	tabela	postgres
public	dopr2003	tabela	postgres
public	dopr2004	tabela	postgres
public	geometry_columns	tabela	postgres
public	parana	tabela	postgres
public	parana_gid_seq	sequência	postgres
public	pop1996	tabela	postgres
public	pop1997	tabela	postgres
public	pop1998	tabela	postgres
public	pop1999	tabela	postgres
public	pop2000	tabela	postgres
public	pop2001	tabela	postgres
public	pop2002	tabela	postgres
public	pop2003	tabela	postgres
public	pop2004	tabela	postgres
public	spatial_ref_sys	tabela	postgres

(30 registros)

#### - Estrutura da tabela *dopr2004*

```
geomedicina=> \d dopr2004
```

Coluna	Tipo	Modificadores
numerodo	character varying(8)	
tipobito	character(1)	
dtobito	character varying(8)	
natural_	character varying(3)	
dt nasc	character varying(8)	
idade	character varying(3)	
sexo	character(1)	
racacor	character(1)	

estciv	character(1)
esc	character(1)
codbaires	character varying(3)
ocup	character varying(5)
codmunres	character varying(7)
lococor	character(1)
codestab	character varying(7)
codmunocor	character varying(7)
idademae	character varying(2)
escmae	character(1)
ocupmae	character varying(5)
qtdfilvivo	character varying(2)
qtdfilmort	character varying(2)
gravidez	character(1)
gestacao	character(1)
parto	character(1)
obitoparto	character(1)
peso	character varying(4)
obitograv	character(1)
obitopuerp	character(1)
assistmed	character(1)
exame	character(1)
cirurgia	character(1)
necropsia	character(1)
causabas	character varying(4)
linhaa	character varying(20)
linhab	character varying(20)
linhac	character varying(20)
linhad	character varying(20)
linhaii	character varying(20)
atestante	character(1)
circobito	character(1)
acidtrab	character(1)
fonte	character(1)
uinform	character varying(2)

geomedicina=>

Em seguida, é necessário realizar uma filtragem para a extração da tabela *dopr2004* do número de óbitos por doença nos municípios. Para isso, criou-se um *script* SQL que seleciona os campos *numerodo*, *codmunres* e *causabas* e faz uma contagem dos óbitos referentes à determinada classe de doença. Uma vez que a classificação da CID-BR agrupa um conjunto de doenças da CID-10, segundo o critério de proximidade do órgão afetado, por exemplo neoplasia maligna do lábio, da cavidade oral e da faringe, doenças que correspondem à classe 032 da CID-BR e às classes C00 a C14 da CID-10, optou-se por usar o recurso de expressões regulares em cláusula *where* as quais conferem um maior poder de expressão que os operadores *like* e *similar to*.

A quantidade de ocorrências de determinada classe de doença é extraída desta tabela e inserida em uma tabela de trabalho, *c00c14*, para posterior cálculo da razão entre esta contagem e o total de habitantes do município naquele ano. O comando usado para esta extração é mostrado a seguir.

```
create table c00c14
as select b.id_ibge, count(a.codmunres)::smallint as qtd
from dopr2004 a right outer join parana b
on a.causabas ~ 'C0[0-9]|C1[0-4]'
and a.codmunres ~ b.id_ibge group by b.id_ibge;
```

A quantidade de mortes no município para cada 100 mil habitantes é atualizada na tabela *c00c14* com o auxílio de consulta à tabela *pop2004*, a qual contém a população do município

no ano de 2004. O comando a seguir calcula a razão e atualiza o campo *tx100mil* com o resultado.

```
geomedicina=> alter table c00c14 add column tx100mil real;
geomedicina=> update c00c14
set tx100mil = (qtd::real/b.pop::real)*100000
from pop2004 b where b.codmun = id_ibge;
```

As razões calculadas são atualizadas no campo *c00c14a2004* da tabela *parana*, a qual é consultada pelo sistema GeoMedicina para a distribuição dos valores em classes do mapa. A seguir é apresentada a estrutura desta tabela.

```
geomedicina=> update parana set c00c14a2004 = (select tx100mil
from c00c14 where parana.id_ibge = c00c14.id_ibge);
```

```
geomedicina=> \d parana
```

Coluna	Tipo	Tabela "public.parana" Modificadores
gid	integer	not null default nextval('public.parana_gid_seq'::text)
id_ibge	character varying	
nome_munic	character varying	
nome_uf	character varying	
meso	bigint	
micro	bigint	
area	numeric	
residentes	bigint	
res_urban	bigint	
res_rural	bigint	
res_0_4	bigint	
res_5_9	bigint	
res_10_19	bigint	
nascidos	bigint	
obitos	bigint	
the_geom	geometry	
taxa_bruta	real	
bayes_loc	real	
bayes_glo	real	
c00c14a2004	real	
c15a2004	real	
c16a2004	real	
c18c21a2004	real	
c22a2004	real	
c25a2004	real	

Índices:

```
"parana_pkey" chave primária, btree (gid)
```

Restrições de checagem:

```
"enforce_dims_the_geom" CHECK (ndims(the_geom) = 2)
```

```
"enforce_geotype_the_geom"
```

```
CHECK (geometrytype(the_geom) = 'MULTIPOLYGON'::text
```

```
OR the_geom IS NULL)
```

```
"enforce_srid_the_geom" CHECK (srid(the_geom) = -1)
```

## B.4 Preenchimento de Tabelas a partir de Arquivos CSV

Em alguns casos, o preenchimento de tabelas é feito a partir de arquivos texto, com campos separados por ponto e vírgula. Esta é uma forma rudimentar de integração de dados entre bases não interconectadas, mas amplamente usada devido à falta de interoperação entre os sistemas. Para esta importação, é necessária a criação de uma tabela na base de dados com campos e tipos que devem casar com os constantes no arquivo CSV, os quais foram obtidos no DATASUS ou SESA e fornecidos pela MINEROPAR. A integridade da importação é garantida

através de chave primária. Exemplos de comandos executados no terminal interativo PSQL para a inserção de dados em tabelas são apresentados a seguir:

#### B.4.1 Preenchimento de Tabela com Número de Casos de Doença

Os arquivos CSV com dados de doenças são obtidos através de uma navegação no site do DATASUS [22], selecionando-se as seguintes opções: INFORMAÇÕES DE SAÚDE, Estatísticas Vitais - Mortalidade e Nascidos Vivos, Mortalidade geral - desde 1979, Abrangência Geográfica: Paraná, Linha: Município, Coluna: Não ativa, Conteúdo: Óbitos p/ Residência; alternando-se entre as opções Períodos Disponíveis: 2004,... Causa - CID-BR-10: Neoplasia maligna do cólon, reto e ânus,...; e escolhendo-se o formato: Colunas separadas por “;”. Outra forma de obtenção dos arquivos CSV é a conversão para este formato de planilhas com dados mais recentes fornecidos pela SESA. Um procedimento para a cópia de dados neste formato é apresentado a seguir.

```
create table hepatite_c (id_munic smallint primary key, pop_munic, num_casos);
copy hepatite_c from '/home/humberto/datasus/hepatite_c.csv' with delimiter ',';
```

#### B.4.2 Preenchimento de Tabela com Teores Geoquímicos

Os arquivos CSV com dados de elementos químicos são obtidos através de uma conversão de planilhas no formato DBF fornecidos pela MINEROPAR. O motivo pelo qual este formato não pode ser transformado diretamente em tabela da base de dados é a formatação das coordenadas de pontos do PostGIS, onde um ponto é expresso na forma POINT(12.45 567.89). Assim, duas colunas do arquivo CSV, com as coordenadas UTM de um ponto da grade regular disposta sobre a área de Estado, têm que ser colocadas em um único campo da tabela PostGIS, denominado geoponto. Para isto foi desenvolvido um *script* em PHP, o qual lê o arquivo em formato CSV e insere o seu conteúdo em uma tabela criada para esta finalidade. Os códigos para a criação da tabela e inserção dos dados na mesma são apresentados a seguir.

Código SQL para a criação da tabela geoquímica

```
create table geoquimica
(
    gid serial not null,
    anocoleta char(4) not null,
```



```

        AL real,
        B real,
        BA real,
        BR real,
        CA real,
        CD real,
        CL real,
        CO real,
        CONDU real,
        CR real,
        CU real,
        F real,
        FE real,
        GA real,
        IN_ real,
        K real,
        LI real,
        MG real,
        MN real,
        MO real,
        NA real,
        NI real,
        NO2 real,
        NO3 real,
        PB real,
        PH real,
        PO4 real,
        SO4 real,
        SR real,
        TL real,
        V real,
        ZN real
    )
;
alter table only geoquimica add constraint geoquimica_pkey primary key (gid, anocoleta);
select AddGeometryColumn ('geoquimica', 'geoponto', 29182, 'POINT', 2);
create index geoquimica_geoponto_idx on geoquimica using gist (geoponto GIST_GEOMETRY_OPS);
VACUUM ANALYSE;

```

### Formulário em HTML para a seleção do arquivo CSV a ser lido

```

<HTML><!-- Formulário para indicação do arquivo CSV -->
<form method='POST' action='learq.php' name='form1' enctype="multipart/form-data">
  <input type='file' name='arq'>
  <input type='submit' value='Ler arquivo'>
</form>
</HTML>

```

### Script em PHP para a leitura do CSV e inserção do seu conteúdo na tabela geoquimica

```

<?php
// conecta à base de dados
$conexao = pg_connect("host=localhost dbname=geomedicina user=humberto");

// lê arquivo CSV e insere os valores na base de dados
$arquivo = $_FILES['arq']['tmp_name'];
$handle = fopen($arquivo, "r") or die("Não conseguiu abrir arquivo");
$delimitador = ',';
$data = fgetcsv ($handle, 1024, $delimitador); //lê o cabeçalho
$row = 1;
while($data = fgetcsv ($handle, 1024, $delimitador))
{
    $num = count ($data);
    if ($num != 34) die("Número de campos diferente do esperado na linha $row");
    $ano = "1996";
    $campo = "anocoleta, geoponto, al, b, ba, br, ca, cd, cl, co, condu, cr, cu, f, fe, ga, in_, k, li, mg, mn, mo, na, ni, no2, no3, pb, ph, po4, so4, sr, tl, v, zn";
    $atrib = "$data[2],$data[3],$data[4],$data[5],$data[6],$data[7],$data[8],$data[9],$data[10], $data[11],$data[12],$data[13],$data[14],$data[15],$data[16],$data[17],$data[18],$data[19],$data[20], $data[21],$data[22],$data[23],$data[24],$data[25],$data[26],$data[27],$data[28],$data[29],$data[30], $data[31],$data[32],$data[33]";
    $coord = "$data[0] $data[1]";
    $query = "insert into geoquimica ($campo) values($ano, GeometryFromText('POINT($coord)', 29182), $atrib)";
    $result = pg_query($conexao, $query);
    if(!$result) die("Erro na inserção SQL");
    $row++;
    echo "Foram inseridos $row registros na tabela geoquimica" . "<br />\n";
}
fclose ($handle);
?>

```

## APÊNDICE C

### CONFIGURAÇÃO DO MAPSERVER

No *MapFile*, o arquivo de configuração do MapServer, é realizada a formatação das diversas camadas que compõem o mapa, bem como a formatação das imagens digitais geradas com a sobreposição destas camadas.

#### C.1 Formatação da Entrada do MapServer - Camadas do Mapa

Cada camada do mapa é definida em um *layer*. No Ka-map é necessário definir grupos para se ter controle sobre a seleção das camadas. As camadas em que não há definição de grupo são agrupadas em um conjunto chamado BASE. No layer são definidos as fontes de dados, no caso, os *ShapeFiles* "bacias.shp", "hg\_solo\_ggrn\_ppb\_poly.shp" e "parana.shp" e a base de dados "parana".

```
# Start of layer definitions
LAYER
  NAME "bacias"
  GROUP "Bacias_hidrográficas"
  METADATA
    "DESCRIPTION" "Bacias Hidrográficas"
  END
  TYPE POLYGON
  STATUS ON
  TRANSPARENCY 40
  DATA 'mineropar/bacias/bacias.shp'
  CLASS
    NAME "Bacias hidrográficas"
    COLOR 240 240 240
    OUTLINECOLOR 199 199 199
  END
END # layer bacias

LAYER
  NAME "hgs"
  GROUP "Hg_no_solo"
  METADATA
    "DESCRIPTION" "Mercúrio no Solo"
  END
  PROJECTION
    "init=epsg:29182"
  END
  TYPE POLYGON
  STATUS OFF
  DATA 'mineropar/elementos/hg_solo_ggrn_ppb_poly.shp'
  CLASSITEM "BASE"
  TRANSPARENCY 70
  CLASS #valor mínimo
    NAME "0.0000 partes por milhão"
    EXPRESSION ([BASE] = 0.0000)
    COLOR 127 127 127
  END
  CLASS #percentil 5
    NAME "29.8700 partes por milhão"
    EXPRESSION ([BASE] = 29.8700)
    COLOR 0 0 0
  END
  CLASS #percentil 15
```

```

        NAME "37.1400 partes por milhão"
        EXPRESSION ([BASE] = 37.1400)
        COLOR 0 0 153
    END
    CLASS #percentil 25
        NAME "42.0600 partes por milhão"
        EXPRESSION ([BASE] = 42.0600)
        COLOR 0 0 128
    END
    CLASS #percentil 40
        NAME "50.4900 partes por milhão"
        EXPRESSION ([BASE] = 50.4900)
        COLOR 0 127 204
    END
    CLASS #percentil 50
        NAME "54.7300 partes por milhão"
        EXPRESSION ([BASE] = 54.7300)
        COLOR 0 191 255
    END
    CLASS #percentil 65
        NAME "62.0000 partes por milhão"
        EXPRESSION ([BASE] = 62.0000)
        COLOR 0 255 255
    END
    CLASS #percentil 75
        NAME "69.0400 partes por milhão"
        EXPRESSION ([BASE] = 69.0400)
        COLOR 0 255 127
    END
    CLASS #percentil 85
        NAME "75.5800 partes por milhão"
        EXPRESSION ([BASE] = 75.5800)
        COLOR 0 128 0
    END
    CLASS #percentil 90
        NAME "80.0600 partes por milhão"
        EXPRESSION ([BASE] = 80.0600)
        COLOR 127 255 0
    END
    CLASS #percentil 91
        NAME "80.9100 partes por milhão"
        EXPRESSION ([BASE] = 80.9100)
        COLOR 255 255 0
    END
    CLASS #percentil 93
        NAME "87.9700 partes por milhão"
        EXPRESSION ([BASE] = 87.9700)
        COLOR 175 141 19
    END
    CLASS #percentil 95
        NAME "105.9200 partes por milhão"
        EXPRESSION ([BASE] = 105.9200)
        COLOR 255 165 0
    END
    CLASS #percentil 97
        NAME "140.5600 partes por milhão"
        EXPRESSION ([BASE] = 140.5600)
        COLOR 255 51 0
    END
    CLASS #percentil 98
        NAME "152.9000 partes por milhão"
        EXPRESSION ([BASE] = 152.9000)
        COLOR 255 0 0
    END
    CLASS #percentil 99
        NAME "161.9700 partes por milhão"
        EXPRESSION ([BASE] = 161.9700)
        COLOR 255 26 0
    END
END # layer hgs

LAYER
    NAME taxa
    GROUP Mortalidade_Infantil
    METADATA
        "DESCRIPTION" "Mortalidade Infantil em 2002"
    END
    TYPE POLYGON
    STATUS OFF
    CONNECTIONTYPE postgis
    CONNECTION "user=postgres dbname=parana host=localhost"
    DATA "the_geom from mortalidade_infantil"
    PROJECTION
        "init=epsg:4291"
    END
    CLASSITEM "taxa_bruta"
    TRANSPARENCY 50
    CLASS
        NAME "Menos que 10 por mil"
        EXPRESSION ([taxa_bruta] < 10)

```

```

        COLOR 255 255 212
    END
    CLASS
        NAME "Entre 10 e 20 por mil"
        EXPRESSION ([taxa_bruta] < 20 AND [taxa_bruta] >= 10)
        COLOR 254 217 142
    END
    CLASS
        NAME "Entre 20 e 30 por mil"
        EXPRESSION ([taxa_bruta] < 30 AND [taxa_bruta] >= 20)
        COLOR 254 153 41
    END
    CLASS
        NAME "Entre 30 e 40 por mil"
        EXPRESSION ([taxa_bruta] < 40 AND [taxa_bruta] >= 30)
        COLOR 217 95 14
    END
    CLASS
        NAME "Mais que 40 por mil"
        EXPRESSION ([taxa_bruta] >= 40)
        COLOR 153 52 4
    END
END # Layer taxa
LAYER
    NAME Parana
    GROUP "Municípios"
    METADATA
        "DESCRIPTION" "Limites municipais"
    END
    TYPE LINE
    MAXSCALE 11600000
    STATUS OFF
    DATA 'ipardes/parana'
    PROJECTION
        "init=epsg:4291"
    END
    CLASSITEM "id_ibge"
    CLASS
        EXPRESSION /.
        OVERLAYSYMBOL 0
        OVERLAYSIZE 1
        OVERLAYOUTLINECOLOR 0 0 0
    END
END # Layer Parana
LAYER
    NAME Parana_nomes
    GROUP "Municípios"
    METADATA
        "DESCRIPTION" "Nomes dos municípios"
    END
    TYPE ANNOTATION
    STATUS OFF
    DATA 'ipardes/parana'
    PROJECTION
        "init=epsg:4291"
    END
    LABELITEM "nome_munic"
    CLASSITEM "nome_munic"
    LABELMAXSCALE 1000001
    CLASS
        EXPRESSION /.
        COLOR -1 -1 -1
        LABEL
            COLOR 0 0 0
            OUTLINECOLOR 202 202 202
            SHADOWCOLOR 202 202 202
            TYPE TRUETYPE
            FONT "vera"
            ANTIALIAS TRUE
            SIZE 8
            POSITION CC
            PARTIALS FALSE
        END
    END # CLASS
    TOLERANCE 1000000
    TOLERANCEUNITS FEET
END # Layer Parana_nomes
LAYER
    NAME "jpl_wms_global_mosaic"
    group "Imagem Satélite NASA"
    TYPE RASTER
    STATUS OFF
    CONNECTION "http://wms.jpl.nasa.gov/wms.cgi?"
    CONNECTIONTYPE WMS
    PROJECTION
        "init=epsg:4326"
    END
    METADATA
        "DESCRIPTION" "JPL Global Mosaic (WMS)"
        "wms_srs" "EPSG:4326"
        "wms_name" "global_mosaic"

```

```

"wms_style" "visual"
"wms_server_version" "1.1.1"
"wms_format" "image/png"
"wms_formatlist" "image/png,image/jpeg,image/gif,image/wbmp"
"wms_connectiontimeout" "120"
"wms_latlonboundingbox" "-54.618972 -26.715458 -48.024250 -22.516105"
"wms_transparent" "true" # para transparencia de imagen
"wms_BGCOLOR" "0xFFFFFFFF"
"wms_boundingbox" "EPSG:4326 -54.618972 -26.715458 -48.024250 -22.516105"
"wms_BBOX" "EPSG:4326 -54.618972 -26.715458 -48.024250 -22.516105"
END # Metadata
END # Layer
END # Map File

```

## C.2 Formatação da Saída do MapServer - Imagem Digital

A imagem digital gerada pelo sistema é controlada pela formatação da saída do MapServer. Nesta formatação consta a extensão (EXTENT) da área total do mapa, o sistema de projeção e o formato da imagem digital. São definidos também os limites mínimo e máximo de escala da saída e o diretório que será acessado pelo servidor HTTP e onde serão gravadas as imagens geradas. Outros elementos visualizados na interface também são configurados nesta seção do *MapFile*, como o mapa de referência, que é uma miniatura de todo o mapa e que serve para referenciar em relação ao todo a parte do mapa que está sendo vista na área da figura, além de permitir a manipulação direta do usuário para a ação de deslocar o mapa; a legenda, que serve para identificar a camada selecionável e semântica das cores ou símbolos aplicados à mesma; a barra de escala que fornece uma indicação das distâncias em relação à ampliação ou redução do mapa.

```

NAME GEOQUIMICA
STATUS ON
SIZE 800 600
SYMBOLSET /var/mapserver/simbolos/symbols.sym
FONTSET /var/mapserver/fontes/fonts.txt
EXTENT 140955.223881 7044256.000000 795044.776119 7506978.723404
UNITS METERS
SHAPEPATH "/home/humberto"
IMAGECOLOR 255 255 255
OUTPUTFORMAT
  NAME png
  DRIVER "GD/PNG"
  MIMETYPE "image/png"
  IMAGEMODE PC256
  EXTENSION "png"
  TRANSPARENT ON
END
DEBUG ON
PROJECTION
  "init=epsg:29182"
# "proj=utm"
# "ellps=GRS67"
# "datum=SAD69"
# "zone=22"
# "units=m"
# "south"
# "no_defs"
END

```

```

#
# Start of web interface definition
#
WEB
  MINSCALE 50000
  MAXSCALE 50000000
  IMAGEPATH "/tmp/kacache/"
  IMAGEURL "/tmp/kacache/"
  LOG "/tmp/geoquimica.log"
END
#
# Start of reference map
#
REFERENCE
  IMAGE /var/mapserver/imagens/pr.png
  EXTENT 140955.223881 7050042.553191 795044.776119 7506978.723404
  STATUS ON
  COLOR -1 -1 -1
  OUTLINECOLOR 255 0 0
  SIZE 210 140
END

#
# Start of legend
#
LEGEND
  TRANSPARENT TRUE
  KEYSIZE 18 12
  LABEL
    TYPE BITMAP
    SIZE MEDIUM
    COLOR 0 0 89
  END
  STATUS ON
END

SCALEBAR
  TRANSPARENT TRUE
  IMAGECOLOR 255 255 255
  LABEL
    COLOR 0 0 0
    SIZE SMALL
  END
  SIZE 150 5
  COLOR 255 255 255
  BACKGROUNDCOLOR 0 0 0
  OUTLINECOLOR 0 0 0
  UNITS kilometers
  INTERVALS 5
  STATUS ON
END

QUERYMAP
  STATUS ON
  STYLE HILITE
  COLOR 255 0 0
END

```